INVESTIGATING ASPARAGINE-LINKED GLYCOSYLATION SUBSTRATE SPECIFICITY AND EFFECTS ON PROTEIN FOLDING

by

Mark M. Chen

B.Sc. Biochemistry & Chemistry
University of British Columbia, 2004

SUBMITTED TO THE DEPARTMENT OF CHEMISTRY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

January 2009
[Feb]

Signature of Author: _____
Department of Chemistry
January 16, 2009

Certified by: _____
Barbara Imperiali
Class of 1922 Professor of Chemistry and Professor of Biology
Thesis Supervisor

Accepted by: _____
Robert W. Field
Haslam and Dewey Professor of Chemistry
Chairman, Departmental Committee on Graduate Students

This doctoral thesis has been examined by a committee of the Department of Chemistry as follows:

Alexander M. Klibanov
Novartis Professor of Chemistry and Bioengineering
Chairman

Barbara Imperiali
Class of 1922 Professor of Chemistry and Professor of Biology
Thesis Supervisor

Catherine L. Drennan
Professor of Chemistry and Professor of Biology

INVESTIGATING ASPARAGINE-LINKED GLYCOSYLATION SUBSTRATE SPECIFICITY AND EFFECTS ON PROTEIN FOLDING

by

Mark M. Chen

Submitted to the Department of Chemistry
on January 16, 2009 in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

Abstract

N-linked glycosylation is a ubiquitous form of protein modification whereby a preassembled oligosaccharide is covalently attached the asparagine side chain of an acceptor protein. This process involves numerous enzymes, produces a diverse set of oligosaccharide structures, and results in a variety of structural and functional effects on the glycoprotein. Research discussed in this dissertation applies synthetic organic chemistry to probe this important biological system.

To study the effects of N-linked glycosylation on protein folding, a semi-synthetic strategy was developed to access a set of model proteins that were homogeneously glycosylated at several sites of interest. The folding kinetics of this set of glycoproteins were then characterized using stopped-flow fluorescence spectroscopy, which revealed that the presence of the glycan show discrete effects on both the rate of protein folding and unfolding, and that the overall effect is highly specific to the local primary and secondary structure of the glycosylation site.

The gram-negative bacterium *Campylobacter jejuni* was recently discovered to contain a general N-linked glycosylation system with a defined glycan structure and tractable enzymes for heterologous expression including a single subunit oligosaccharyltransferase. To probe the bacterial N-linked glycosylation machinery, a chemo-enzymatic synthesis for each of the glycan intermediates within this pathway was developed, which are impractical to obtain from the host organism. Importantly, chemo-enzymatic allowed for the incorporation of structural modifications for binding-specificity assays and radiolabels for accurate quantification. Access to these substrates allowed us to define the minimum glycosylation consensus sequence for the oligosaccharyltransferase as well as the polyisoprenol specificity of three representative enzymes within the pathway.

Thesis Supervisor: Barbara Imperiali
Class of 1922 Professor of Chemistry and Professor of Biology

Acknowledgements

First and foremost I am grateful beyond adequate acknowledgement to my supervisor Barbara Imperiali for her guidance, support, and mentorship throughout my graduate studies. She is a role model in many ways, and I have tried my best to emulate her approach and attitude towards research. I admire her especially for her enthusiastic and courageous pursuit of difficult problems in science, her ability to clearly and eloquently explain scientific concepts, and her steadfast attention to laboratory safety and the well being of each lab member. I consider myself extremely fortunate to have been a graduate student under her supervision and to have conducted research in such a positive and supportive environment.

I would like to thank the members of my thesis committee, especially professors Alexander Klibanov and Catherine Drennan, as well as Joanne Stubbe and Stuart Licht for their advice, feedback, and encouragements. The second year orals and third year research proposal were both extremely useful in helping me recognize my abilities and shortcomings, and how I can improve myself as a scientist. I am grateful for their willingness to share their time and experience.

I would like to thank my lab mates who have helped me in the preparation of this dissertation, including Angelyn Larkin, Cliff Stains, Elvedin Lukovic, Galen Loving, James Morrison, Jay Troutman, Michael Morrison, Marcie Jaffee, Matthieu Sainlos, Meredith Hartley, Michelle Chang, and Wendy Iskenderian. I am grateful for their numerous insightful comments and thoughtful suggestions.

Barbara often says "it takes a forest to grow a tree", and I must concur that I could not have made it through graduate school without the help and support of my wonderful lab mates. I am especially grateful to my pod mates Langdon Martin, Guofeng Zhang, Meredith Hartley and Cliff Stains. I could not have asked for a more wonderful company to sit with even if I had handpicked them myself. I hold Langdon in the highest esteem, for the way he maintains a balance between research and everything else. He played an invaluable and often thankless role in the lab by organizing volleyball and sushi, writing witty birthday emails, doing more than his fair share of lab cleanup and chores, and providing quirky facts and helpful references. Likewise I have always admired Meredith for her approach to the ups and downs of graduate school. Exams and minimeetings never seem to stress her out, and neither has her difficult research. She brings a cheerful and positive vibe to the lab, which always helps to get me through my times of anxiety.

I would like to thank Jay Troutman for all the late nights in the lab we spent doing our organic extraction assays together. I am grateful to have someone to discuss the finer points of enzyme kinetics as well as someone to commiserate with over the long hours necessary to carry out these assays. I thank Nelson Olivier for being a wonderful instructor of various biochemical techniques. I admire his approach to research: always focused and deliberate, and never taking shortcuts. I thank Galen Loving for welcoming me to the lab when I first came to MIT and for the many scientific and philosophical discussions we have had. He is one deep thinker and well-read individual and I appreciate his many thoughtful and insightful views. I thank Elvedin

Lukovic and Angelyn Larkin for making the Imperiali lab and the department in general a lot of fun. They are arguably the two of the most social and witty individuals that I have met in graduate school, and whether it is at the Muddy or at a restaurant, I always seem to have the best time when I am seated near them. I thank Matthieu Sainlos and Brenda Goguen for the pleasure of their company working the "late shift". They are two of the hardest working people that I know, and their work ethic has always been both a source of motivation and inspiration. I thank Wendy Iskenderian for keeping the lab safe and energy efficient. I thank Eranthie Weerapana, Jebrell Glover, Guofeng Zhang, Mary O'Reilly and Christian Hackenberger for making much of my dissertation possible through the important groundwork they carried out, as well as for introducing me to many new techniques and protocols.

I thank Debby Pheasant for all her help with my experiments in the BIF. She is one of the most helpful and dedicated staff I have ever met, and she does a wonderful job of keeping so many instruments all running smoothly. I cannot thank Elizabeth Fong enough for being so extremely organized and for the countless occasions where she helped me with appointments, bookings, mailings and saved me from my last-minute planning. I thank Susan Brighton for welcoming me to the department and for taking care of me every since. I have met very few administrators who show such dedication to their students and know each of us by name. I thank Juan González-Vera, Anne Reynolds, Katja Barthelmes, Melanie Bonnekessel, Andreas Aemissegger, Dora Carrico-Moniz, Elizabeth Vogel, Seungjib Choi, and Bianca Sculimbrene for making the Imperiali lab such a supportive and pleasant place to work. I wish James Morrison all the best as he fulfills the role of resident Canadian in the lab. I am glad it is someone as smart, hardworking, and classy as this fellow UBC Alum. To the new members of the lab Marcie Jaffee, Cliff Stains, Michael Morrison, and Michelle Chang, I sincerely wish each of you all the success on your exciting new projects and I hope you will enjoy working here as much as I did. I would also like to thank everyone from the Stubbe lab for being such wonderful neighbors and for providing a generous and invaluable source of expertise, equipment, and friendship.

I have always loved research, but when research sometimes does not cooperate I am grateful to be surrounded by friends and family to keep me focused and on-track. I would like to give a shout out to my workout partner Sean Smith for helping me stay in shape, Leonardo Gomez and Dustin Siegel for being such great room mates, Allan Fong for introducing me to his bible study group and the best volleyball team I have ever played with, Shuo Zhang and Ran Yi and everyone from CSSA Volleyball team for replacing tennis with volleyball as my favorite summer time sport, everyone at the SP bible study group for a lot of great food and interesting discussions every week, my tennis partner Tim Chan for career advice, and Jiamin Chin for teaching me so much about myself and how I can be a better person.

I would like to thank everyone back home in Vancouver for keeping in touch and for helping me "recharge" every time I come back, especially my buddy Eugene Choo for all the mountains we skied or hiked, Patrick and Stephanie Chen for the many years of tennis, golf, and now skiing, Robert An and Jonny Lee and the rest of the fam for all the fun times and rounds of centurions, Dennis Wang for all the volleyball and hold'em and for showing me a great time while in the UK, my mentors Stephen Withers and Paul Stinson for always providing me with great advice and encouragements, my cousins Hattie Chen, Linda Jung, and Gloria Chen for being such wonderful role-models to look up to when I was growing up, and countless others (Simon Chow,

Christine Parachoniack, Daphne Shih, Erina Chan, Alexandra Tan, Carolyn Yeung, Able Fok, Andrea De Souza, Joey Lin, Eric Hsih, Michael Eastwood, Eagranie Yuh, Leo Lai, Brendon Luc, Albert Tam, Alice Wang, Anthony Nguyen, and Jenni Chen) for always making me feel blessed to have such wonderful and loyal friends. I would also like to thank all my family around the world, especially my grandparents, Uncle DayDay, Uncle Edward and Aunt Ruth, Uncle Stanley and Aunt Vivian, Uncle Henry, Greatuncle Wilfred and Greataunt Minowei for their love and support.

And finally I would like to thank my parents David Chen and Lucy Yang. I respect them more than anyone in the world and I have to acknowledge that every major decision or accomplishment I have ever made has been strongly guided by their advice and encouragement. I have always been motivated by their courage, conviction, and hard work as new immigrants to Canada. They have taught me both by words and by example, and I attribute any and all my success to their positive influence.

Dedicated to my wonderful parents
David and Lucy

Table of Contents

List of Figures

Chapter 4 – Chemo-Enzymatic Synthesis of Undecaprenol-Linked Substrates

Chapter 5 – Oligosaccharyltransferase PglB Acceptor Specificity

Chapter 6 – Polyisoprene Specificity in the Pgl Pathway

List of Tables

List of Abbreviations

Ac	acetyl
AcCoA	acetyl coenzyme A
All	allyl
ATP	adenosine 5'-triphosphate
bacillosamine	2,4,6-trideoxy-2,4-diacetamido-D-glucopyranose
bacillosamine(6OH)	2,4-dideoxy-2-4-diacetamido-D-glucopyranose
Bn	benzyl
Bz	benzyol
C. jejuni	*Campylobacter jejuni*
CBD	chitin binding domain
CD	circular dichroism
CE	capillary electrophoresis
CNX	calnexin
CRT	calreticulin
Dap	2,3-diaminopropionic acid
DIPEA	diisopropylethylamine
DMAP	4-dimethylaminopyridine
DMF	dimethylformamide
DMSO	dimethylsulfoxide
DPM	disintegrations per minute
DTT	dithiothreitol
E. coli	*Escherichia coli*
ε	extinction coefficient or molar absorptivity
EDT	1,2-ethanedithiol
EDTA	ethylenediaminetetraacetic acid
EPL	expressed protein ligation
ER	endoplasmic reticulum
ESI-MS	electrospray ionization mass spectrometry
Fmoc	fluoren-9-ylmethyloxycarbonyl
GalNAc	*N*-acetyl-D-galactosamine
GDP	guanosine 5'-diphosphate
GlcNAc	*N*-acetyl-D-glucosamine
Glc	D-glucose
HBTU	2-(1*H*-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HEPES	*N*-(2-hydroxyethyl)-piperazine-*N*'-(2-ethane-sulfonic acid)
HOBt	hydroxybenzotriazole
HPLC	high-performance liquid chromatography
Hse	homoserine

Im7	immunity protein 7
IPTG	isopropyl-β-D-thiogalactopyranoside
LB	Luria-Bertani
Man	D-mannose
MeCN	acetonitrile
MESNA	sodium mercaptoethane sulfonate
N. gonorrhoeae	*Neisseria gonorrhoeae*
NAD^+	nicotinamide adenine dinucleotide
NMP	1-methyl-2-pyrrolidinone
NMR	nuclear magnetic resonance
OD_{600}	optical density at 600 nm
OT	oligosaccharyltransferase
PAL	5-(4'-aminomethyl-3',5'-dimethoxyphenoxy)valeric acid
PEG-PS	polyethyleneglycol-grafted polystyrene
Pgl pathway	protein glycosylation pathway (in *Campylobacter jejuni*)
PLP	pyridoxal-5'-phosphate
*p*NF	*p*-nitrophenylalanine
PSUP	pure solvent upper phase
PyAOP	azabenzotriazol-1-yloxy-tris(pyrrolidino)phosphonium hexafluorophosphate
PyBOP	benzotriazole-1-yl-oxy-tris(pyrrolidino)phosphonium hexafluorophosphate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
S. cerevisiae	*Saccharomyces cerevisiae*
SBA	soybean agglutinin
SPPS	solid-phase peptide synthesis
SRP	signal recognition particle
TBDMS	tert-butyldimethysilyltrifluoromethane
TBDMS-OTf	tert-butyldimethysilyltrifluoromethane sulfonate
TFA	trifluoroacetic acid
THF	tetrahydrofuran
TNBS	2,4,6-trinitrobenzenesulfonic acid
TLC	thin layer chromatography
Tris	tris(hydroxymethyl)aminomethane
Ts	tosyl
TUP	theoretical upper phase
UDP	uridine 5'-diphosphate
UV-Vis	ultraviolet-visible

CHAPTER 1

INTRODUCTION

N-linked glycosylation is an enzymatic process that takes place in eukarya, archaea, as well as bacteria domains (1), where a preassembled oligosaccharide is covalently attached to the asparagine side-chain nitrogen of proteins. Within eukaryotes, N-linked glycosylation has been estimated to occur on approximately 50% of all proteins (2). Like phosphorylation, this common form of protein modification can have dramatic effects on protein structure and function. Unlike phosphorylation, which generally imparts an on/off effect, N-linked glycans allow for many downstream effects due to the diversity of glycans employed (**Figure 1-1**) (3,4). The diversity of glycan structure and the extensive enzyme machinery that is needed to produce this diversity makes N-linked glycosylation and N-linked glycoproteins especially challenging and interesting to study. This introduction will provide a brief background on N-linked glycosylation as it occurs in eukaryotes and the bacterium *Campylobacter jejuni*, and its general effects on glycoproteins.

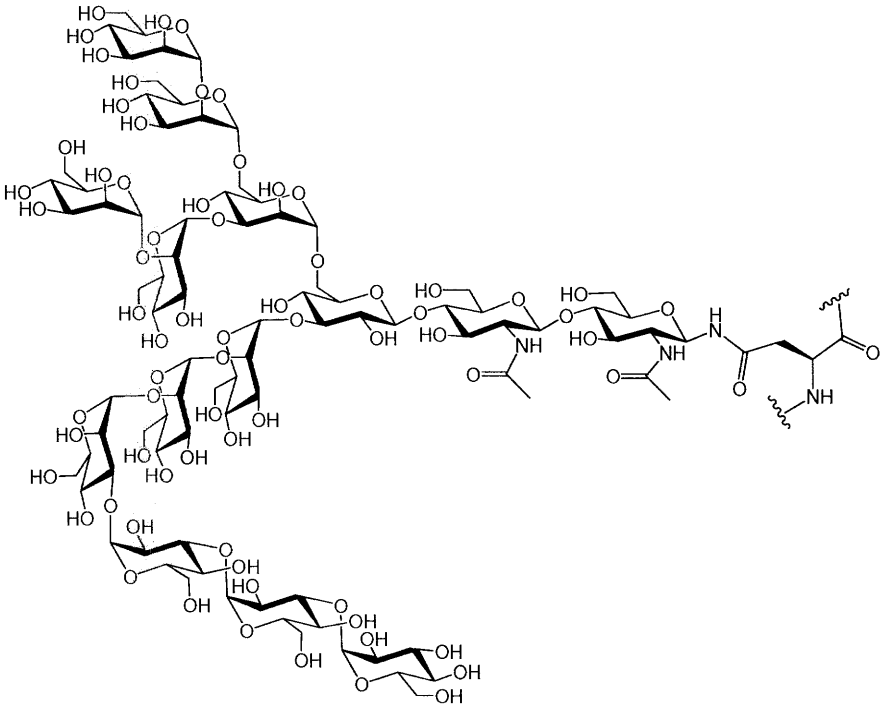


Figure 1-1. Eukaryotic N-linked glycan in *S. cerevisiae* attached to the asparagine side chain of the acceptor protein (5). Wavy lines represent the extension of the polypeptide.

1-1. N-linked glycosylation in eukaryotes

Eukaryotic N-linked glycosylation has been most extensively studied in the yeast system of *Saccharomyces cerevisiae* (3,6). The process begins on the cytosolic surface of the endoplasmic reticulum (ER) where a series of glycosyltransferases build up an oligosaccharide on a dolichyl diphosphate-carrier using uridine diphosphate (UDP) and guanosine diphosphate (GDP) activated sugars (**Figure 1-2**). Dolichols are a family of α-saturated polyisoprenols consisting of 14-17 isoprene units, which is embedded within the membrane bilayer (7). The glycan is assembled to the point of a heptasaccharide of defined structure, consisting of two core N-acetylglucosamine (GlcNAc) and five mannose (Man) residues in a branched structure (3).

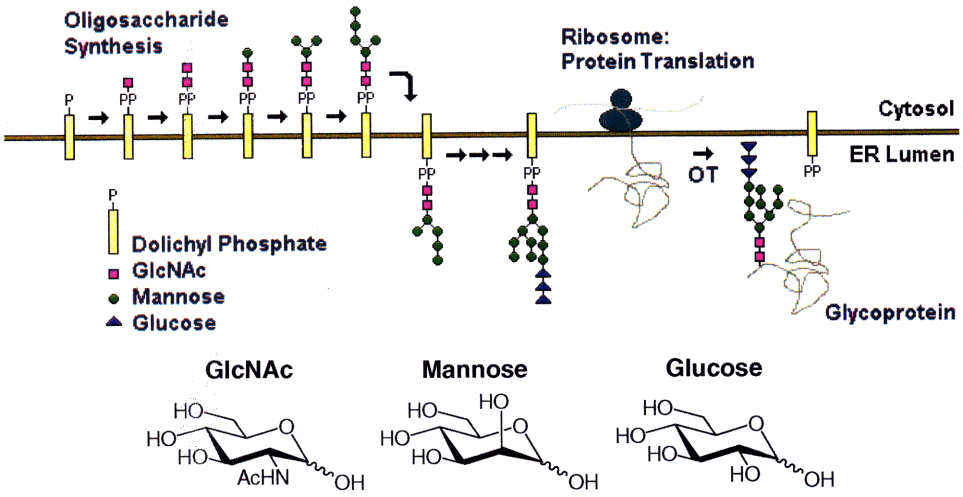


Figure 1-2. The eukaryotic N-linked glycosylation pathway in *S. cerevisiae*. Following its transfer to the acceptor protein, the tetradecasaccharide glycan undergoes additional processing in the ER and the Golgi apparatus. GlcNAc = N-acetylglucosamine.

After this heptasaccharide is assembled, it is flipped from the cytosolic to the lumenal face of the ER membrane by an ATP-independent, bi-directional, membrane-spanning flippase (8). Within the ER, glycan assembly is continued by another series of glycosyltransferases, which work

sequentially to add on four additional branching Mans and three terminal glucoses (Glc) to form the tetradecasaccharide "core" unit (**Figure 1-1**) (3). One important difference in this case is that instead of using water soluble nucleotide-activated UDP-sugar or GDP-sugar donors, the glycosyltransferases within the ER lumen use membrane-associated, dolichyl phosphate-bound sugar donors.

Transfer of the tetradecasaccharide to protein occurs in the ER lumen. Potential glycoproteins that are designated for the secretory pathway contain a signal sequence that is recognized by the signal recognition particle (SRP), which directs the sequence to the translocation machinery for transport across the ER membrane (9,10). The signal sequence is then cleaved by a dedicated protease (11). At this point, the nascent protein interacts with the oligosaccharyltransferase (OT), which takes the fully assembled tetradecasaccharide from the dolichyl diphosphate-carrier and transfers it to the asparagine side chain of an acceptor protein within the consensus sequence Asn-X-Ser/Thr where X can be any amino acid except for proline (12,13). This process occurs as protein translation is taking place. Following this transfer, the glycan undergoes further processing throughout the ER and the Golgi apparatus where additional glycosyltransferases and glycosidases modify the glycan structure (4,14).

The *S. cervisiae* OT consists of at least eight membrane bound protein subunits, five of which are essential for yeast viability (**Figure 1-3**) (12,13). Each of the subunits is highly hydrophobic and many are themselves modified by N-linked glycans. This protein complex cannot be effectively recombinantly overexpressed in *Escherichia coli* or in baculovirus-infected insect cells, and has been found to be extremely difficult to purify. As a result, precise biochemical and biophysical

characterizations are very difficult to carry out. The chemistry that the OT catalyzes is quite remarkable because it is capable of activating the nitrogen of the asparagine amide side chain which is an extremely poor nucleophile. Currently, the molecular details of this reaction mechanism remain uncertain. One possible mechanism involves the formation of an Asx-turn, which would help protonate the carbonyl oxygen and allow an enzyme-mediated deprotonation of the nitrogen to tautomerize the amide into a neutral nucleophilic imidol species (**Figure 1-4**).

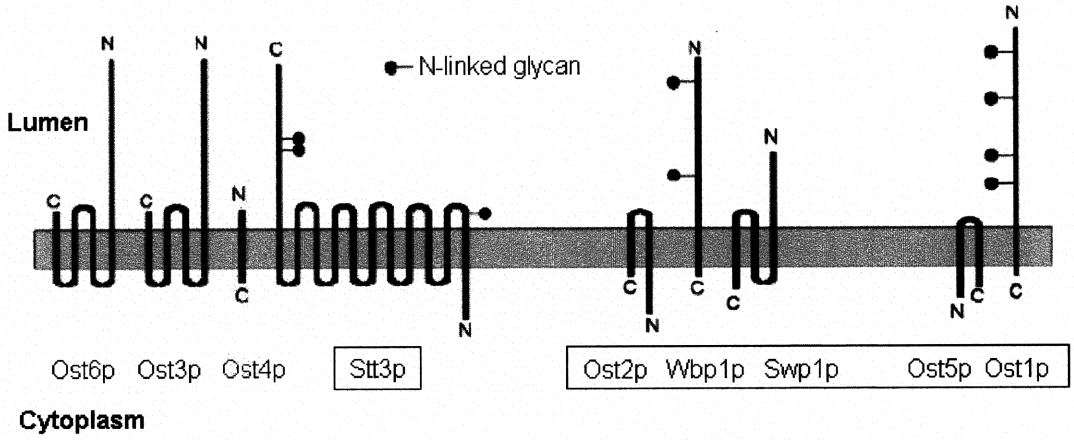


Figure 1-3. Subunit composition of the *S. cerveisiae* oligosaccharyltransferase complex (6). The boxed subunits are essential for yeast viability. Only one of Ost6p or Ost3p is present in a single OT complex.

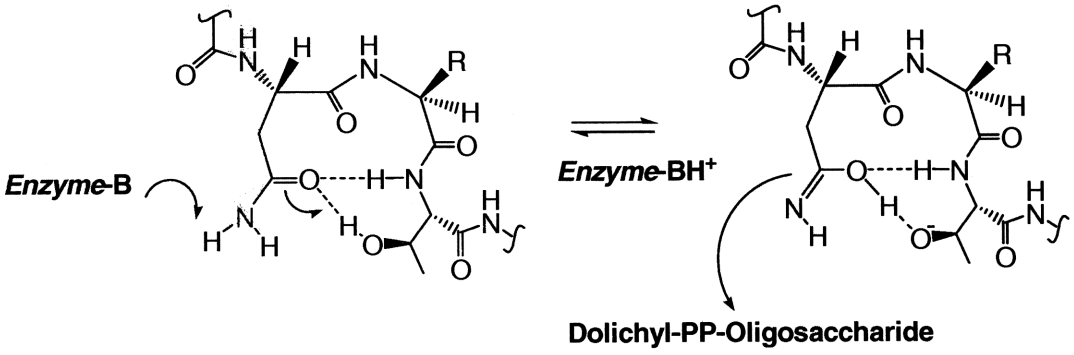


Figure 1-4. Proposed mechanism of the oligosaccharyltransferase.

1-2. N-linked glycosylation in *Campylobacter jejuni*

For many years, N-linked glycosylation was thought to be unique to the eukarya and archaea domains (1,6,15). Recently, *C. jejuni* became the first bacterium found to contain a general N-linked protein glycosylation pathway (Pgl) (16,17). Although the enzymes and substrates are different, the overall process shares many parallels with eukaryotic pathway in *S. cerevisiae*. The Pgl pathway consists of a total of nine enzymes (**Figure 1-5**) (18). Three enzymes, PglF, PglE, and PglD, function to biosynthesize the UDP-activated form of an unusually modified sugar, 2,4-diacetamido-bacillosamine (bacillosamine) at the cytosolic face of the periplasmic membrane. These three enzymes function as a dehydratase, an aminotransferase, and an acetyltransferase, respectively, creating the necessary modifications to the UDP-N-acetylglucosamine (UDP-GlcNAc) starting material.

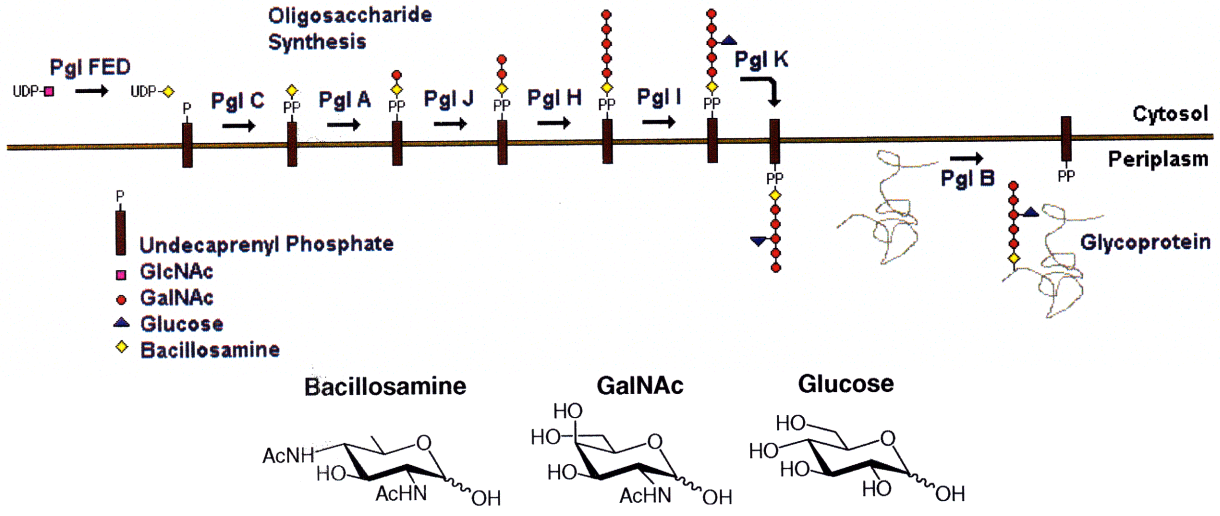


Figure 1-5. The prokaryotic N-linked glycosylation pathway in *C. jejuni*. GalNAc = N-acetylgalacosamine; GlcNAc = N-acetylglucosamine; bacillosamine = 2,4-diacetamido-bacillosamine.

Bacillosamine is the first sugar that is transferred to the polyisoprenol-carrier (18,19). The carrier in this case is undecaprenol, an α-unsaturated polyisoprenol consisting of 11 isoprene units, which is shorter than the eukaryotic dolichol, but still very hydrophobic and embedded within the membrane. PglC is the glycosylphosphotransferase that transfers bacillosamine phosphate onto undecaprenyl phosphate (19). This is followed by the actions of four more glycosyltransferases, PglA, PglJ, PglH, and PglI which sequentially add five additional N-acetyl-galactosamines (GalNAc) and one branching Glc to form the full heptasaccharide (**Figure 1-6**) (20). Similar to the eukaryotic system, the polyisoprenyl diphosphate-linked heptasaccharide is flipped by a flippase, PglK, across the membrane into the periplasm (21). Within the periplasm, the heptasaccharide is transferred by the *C. jejuni* OT, PglB, from the undecaprenol-carrier to the asparagine side chain of the acceptor protein (22,23). Unlike the eukaryotic system, no further processing of the heptasaccharide is known to occur after it has been transferred to the protein acceptor.

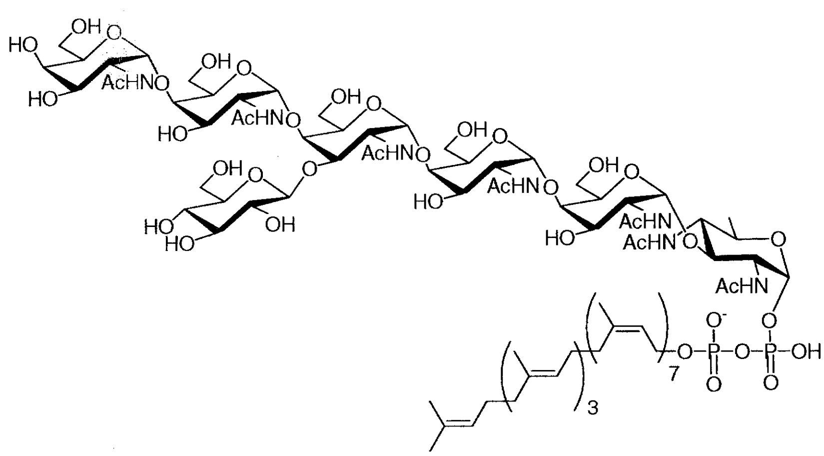


Figure 1-6. *C. jejuni* N-linked glycan donor: undecaprenyl diphosphate-heptasaccharide.

Another major difference between *C. jejuni* pathway and the prototypic eukaryotic counterpart is found in the OT structure. In *C. jejuni*, PglB consists of only a single membrane bound subunit

as opposed to the eight subunits in the eukaryotic complex. However, PglB shares sequence homology and membrane topology with the Stt3p subunit of the *S. cerevisiae* OT, suggesting that they have a similar function and are likely to operate by the same mechanism (22). Advantageously, PglB and the other enzymes in the Pgl pathway can be recombinantly overexpressed in *E. coli* (23). This makes PglB a much more promising target for mechanistic and crystallographic studies than the eukaryotic OT. Although many archaeal systems also possess single subunit OT, limited information is known about their glycan structure and synthetic pathway (15). The Pgl pathway of *C. jejuni* overall provides a very useful model for N-linked glycosylation because it has accessible substrates, a smaller number of enzymes, and is relatively straightforward to express and purify.

1-3. The general effects of N-linked glycosylation

It is important to note that the N-linked glycan is fairly large in comparison to the protein to which it is attached, especially considering that many proteins contain multiple glycosylation sites. Each glycan may weigh up to 3 kDa and can extend 3 nm or more into solution (4). These glycans are flexible, hydrated branched structures that may behave relatively independently of the protein. The presence of this large hydrophilic appendages can have many different effects, depending upon its location on the protein (4). N-linked glycans have been shown to improve protein solubility by covering exposed hydrophobic patches on protein surfaces (24), and to increase proteolytic stability by sterically blocking protease access (25). They have also been known to modulate enzyme activity although the exact cause of this effect has not generalized (26).

N-linked glycans are ideal tags for signaling because the chemistry of their glycosidic linkages allows for incredible diversity and versatility in their structure, providing a large amount of information within a localized space (4,27). Glycan are known to interact with many carbohydrate-binding proteins called lectins through weak but cooperative binding. Many of these lectins are involved in protein transport (28), antigen recognition (29), or as chaperones for protein folding (30). Recognition of the N-linked oligosaccharide allows lectins to determine protein localization, specific cell types, and whether it is properly folded. As the glycans are processed in the ER and the Golgi complex, subtle modifications to the glycan structure allows for adjustments in these interactions.

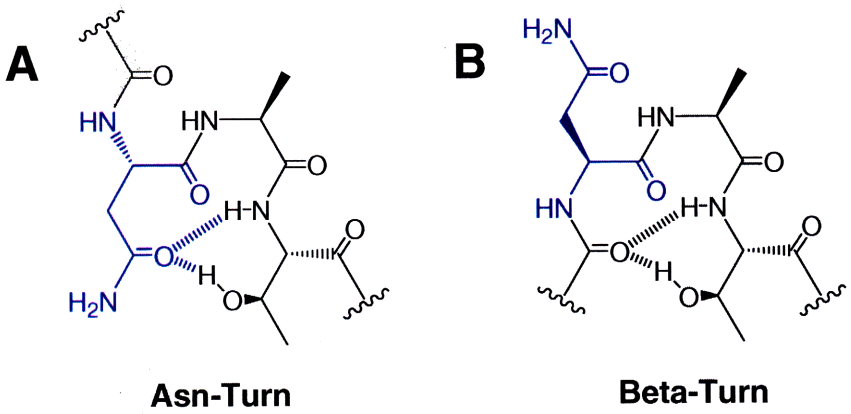


Figure 1-7. Example of **A**) an Asx-turn and **B**) a β-turn. The curvy lines represent the extension of the polypeptide.

In vitro and in silico studies have also revealed the ways in which the N-linked glycan can more directly impact protein structure (31). Studies on glycopeptides reveal a dramatic conformational change upon glycosylation from an extended Asx-turn to a more compact type 1 β-turn (**Figure 1-7**) (32). Surprisingly this structural change does not appear to arise from hydrogen-bond interactions between the glycan and the peptide, but rather through steric

crowding or solvent-mediated effects (5). Furthermore, the effects of the glycan are heavily dependent on the specific structure of the glycan. Substitution of the inner two GlcNAcs for glucoses (33), and replacement of the native β-asparagine linkage to an α-linkage (34) both fail to induce β-turn formation. It has been proposed that the presence of an N-linked glycan in critical locations within the protein may promote protein folding by restricting the dihedral angles and the overall conformational space available to the local structure (31,35).

1-4. Dissertation objectives

Eukaryotic N-linked glycoproteins typically contain glycans of very heterogeneous structures because of the many glycosyltransferases and glycosidases that are involved in their processing (3). This structural diversity makes it very difficult to obtain homogeneously glycosylated protein for biophysical studies. In Chapter 1, we describe a semi-synthetic approach to obtaining N-linked glycoproteins with direct control over the glycan structure and site. Using this approach we were able to synthesize a set of highly pure and homogeneously-glycosylated proteins in milligram quantities.

Since N-linked glycosylation is a co-translational process in eukaryotic organisms, and occurs before the protein has reached its native structure, it has long been implicated to have an effect on protein folding (6). Many studies on glycopeptides have already shown that the presence of the N-linked glycan can dramatically effect the peptide conformation (32). In Chapter 2, we describe quantitative folding kinetic studies using the glycoproteins synthesized from the previous chapter. We found that the presence of the glycan can result in distinct changes to both

the rate of protein folding and unfolding, and that the effect is highly specific to the local structure of the glycosylation site.

Bacillosamine is a highly modified sugar in the *C. jejuni* N-linked glycosylation pathway (36). As the first sugar transferred to the polyisoprenyl diphosphate-carrier, bacillosamine is part of the substrates of every glycosyltransferase and the OT in the Pgl pathway. Since bacillosamine is not commercially available and cannot be obtained in sufficient quantities within this pathogenic bacterium, we describe in Chapter 3 a chemo-enzymatic synthesis to obtain this unusual sugar. Access to large quantities of highly pure bacillosamine and its derivatives (**Figure 1-8**) greatly facilitated our ability to probe many of the important enzymes within the pathway.

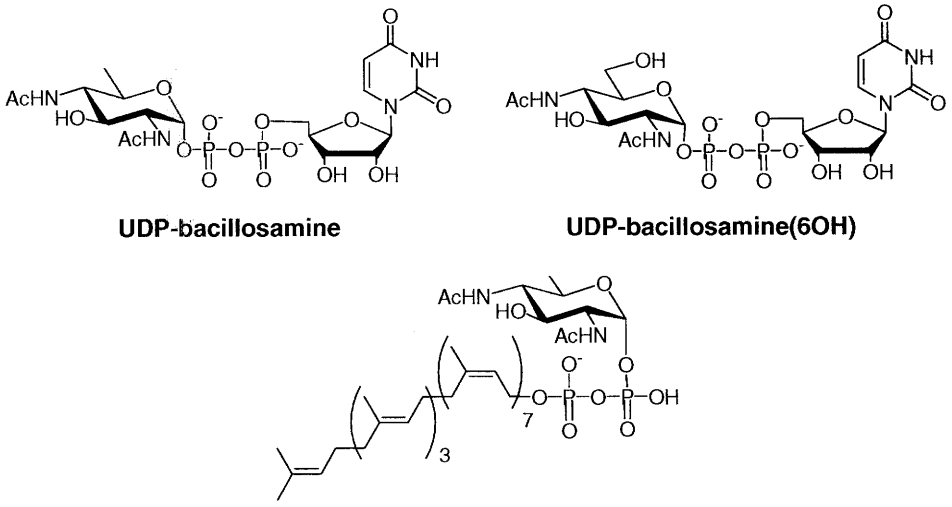


Figure 1-8. Structure of bacillosamine derivatives used in the study of the Pgl pathway.

In eukaryotic N-linked glycosylation, the OT transfers the glycan to the asparagine side chain of the sequence Asn-X-Ser/Thr where X can be any amino acid other than proline (37). Using a small library of peptides and a radiolabeled undecaprenol-linked glycan substrate made from our newly synthesized UDP-bacillosamine, we describe in Chapter 4 the determination of the minimum consensus glycosylation sequence for the *C. jejuni* OT, PglB. Interestingly, PglB appears to contain additional sequence requirements beyond the three residue motif. By systematically altering the amino acids at each residue, we were also able to determine the optimal sequence for PglB.

The interaction between membrane-associated enzymes and their lipophilic substrates has not been well understood. The polyisoprene can either play a simple physical role in anchoring the glycan to the membrane where the enzymes act, or it may play a more specific chemical role in substrate recognition. In Chapter 5, we describe the incorporation bacillosamine into substrates with non-native polyisoprenols and the assays performed using these non-native substrates with three enzymes from the *C. jejuni* pathway, each with different membrane topology. The non-native polyisoprenols differed in their overall length, degree of unsaturation, and double bond geometry. Probing the enzymes with these substrate analogues allowed us to derive insight into the two possible roles for the polyisoprenes.

References

1. Spiro RG (2002) Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **12**: 43R-56R.

2. Apweiler R, Hermjakob H, Sharon N (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta* **1473**: 4-8.

3. Burda P, Aebi M (1999) The dolichol pathway of N-linked glycosylation. *Biochim. Biophys. Acta.* **1426**: 239-57.

4. Helenius A, Aebi M (2004) Roles of N-linked glycans in the endoplasmic reticulum. *Annu. Rev. Biochem.* **73**: 1019-49.

5. Imperiali B, O'Connor SE (1999) Effect of N-linked glycosylation on glycopeptide and glycoprotein structure. *Curr. Opin. Chem. Biol.* **3**: 643-9.

6. Weerapana E, Imperiali B (2006) Asparagine-linked protein glycosylation: From eukaryotic to prokaryotic systems. *Glycobiology* **16**: 91-101.

7. Schenk B, Fernandez F, Waechter CJ (2001) The ins(ide) and out(side) of dolichyl phosphate biosynthesis and recycling in the endoplasmic reticulum. *Glycobiology* **11**: 61R-70R.

8. Helenius J, Ng DT, Marolda CL, Walter P, Valvano MA, Aebi M (2002) Translocation of lipid-linked oligosaccharides across the ER membrane requires Rft1 protein. *Nature* **415**: 447-50.

9. Halic M, Beckmann R (2005) The signal recognition particle and its interactions during protein targeting. *Curr Opin Struct Biol* **15**: 116-25.

10. Shan SO, Walter P (2005) Co-translational protein targeting by the signal recognition particle. *FEBS Lett* **579**: 921-6.

11. Bohni PC, Deshaies RJ, Schekman RW (1988) SEC11 is required for signal peptide processing and yeast cell growth. *J Cell Biol* **106**: 1035-42.

12. Kelleher DJ, Gilmore R (2006) An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology* **16**: 47R-62R.

13. Lennarz WJ (2007) Studies on oligosaccharyl transferase in yeast. *Acta Biochim Pol* **54**: 673-7.

14. Munro S (2001) What can yeast tell us about N-linked glycosylation in the Golgi apparatus? *FEBS Lett* **498**: 223-7.

15. Messner P (2004) Prokaryotic glycoproteins: unexplored but important. *J Bacteriol* **186**: 2517-9.

16. Szymanski CM, Logan SM, Linton D, Wren BW (2003) *Campylobacter*--a tale of two protein glycosylation systems. *Trends. Microbiol.* **11**: 233-8.

17. Szymanski CM, Wren BW (2005) Protein glycosylation in bacterial mucosal pathogens. *Nat. Rev. Microbiol.* **3**: 225-37.

18. Linton D, Dorrell N, Hitchen PG, Amber S, Karlyshev AV, Morris HR, Dell A, Valvano MA, Aebi M, Wren BW (2005) Functional analysis of the *Campylobacter jejuni* N-linked protein glycosylation pathway. *Mol. Microbiol.* **55**: 1695-703.

19. Glover KJ, Weerapana E, Chen MM, Imperiali B (2006) Direct biochemical evidence for the utilization of UDP-bacillosamine by PglC, an essential glycosyl-1-phosphate transferase in the *Campylobacter jejuni* N-linked glycosylation pathway. *Biochemistry* **45**: 5343-50.

20. Glover KJ, Weerapana E, Imperiali B (2005) *In vitro* assembly of the undecaprenylpyrophosphate-linked heptasaccharide for prokaryotic N-linked glycosylation. *Proc. Natl. Acad. Sci. U.S.A.* **102**: 14255-9.

21. Alaimo C, Catrein I, Morf L, Marolda CL, Callewaert N, Valvano MA, Feldman MF, Aebi M (2006) Two distinct but interchangeable mechanisms for flipping of lipid-linked oligosaccharides. *EMBO J.* **25**: 967-76.

22. Glover KJ, Weerapana E, Numao S, Imperiali B (2005) Chemoenzymatic synthesis of glycopeptides with PglB, a bacterial oligosaccharyl transferase from *Campylobacter jejuni*. *Chem. Biol.* **12**: 1311-5.

23. Wacker M, Linton D, Hitchen PG, Nita-Lazar M, Haslam SM, North SJ, Panico M, Morris HR, Dell A, Wren BW, Aebi M (2002) N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science* **298**: 1790-3.

24. Hoiberg-Nielsen RR, Arleth L, Westh P (2008) The effect of glycosylation on interparticle interactions and dimensions of native and denatured phytase. *Biophys J*

25. Kundra R, Kornfeld S (1999) Asparagine-linked oligosaccharides protect Lamp-1 and Lamp-2 from intracellular proteolysis. *J Biol Chem* **274**: 31039-46.

26. Stengel C, Newman SP, Day JM, Tutill HJ, Reed MJ, Purohit A (2008) Effects of mutations and glycosylations on STS activity: a site-directed mutagenesis study. *Mol Cell Endocrinol* **283**: 76-82.

27. Helenius A, Aebi M (2001) Intracellular functions of N-linked glycans. *Science* **291**: 2364-9.

28. Li JG, Chen C, Liu-Chen LY (2007) N-Glycosylation of the human kappa opioid receptor enhances its stability but slows its trafficking along the biosynthesis pathway. *Biochemistry* **46**: 10960-70.

29. Imberty A, Varrot A (2008) Microbial recognition of human cell surface glycoconjugates. *Curr Opin Struct Biol* **18**: 567-76.

30. Caramelo JJ, Parodi AJ (2008) Getting in and out from calnexin/calreticulin cycles. *J Biol Chem* **283**: 10221-5.

31. Meyer B, Moller H (2007) Conformation of glycopeptides and glycoproteins. *Top. Curr. Chem.* **267**: 187-251.

32. O'Connor SE, Imperiali B (1997) Conformational Switching by Asparagine-Linked Glycosylation. *J. Am. Chem. Soc.* **119**: 2295-2296.

33. O'Connor SE, Imperiali B (1998) A molecular basis for glycosylation induced conformational switching. *Chem. Biol.* **5**: 427-437.

34. Bosques CJ, Tschampel SM, Woods RJ, Imperiali B (2004) Effects of glycosylation on peptide conformation: a synergistic experimental and computational study. *J Am Chem Soc* **126**: 8421-5.

35. Petrescu AJ, Milac AL, Petrescu SM, Dwek RA, Wormald MR (2004) Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology* **14**: 103-14.

36. Sharon N (2007) Celebrating the golden anniversary of the discovery of bacillosamine, the diamino sugar of a Bacillus. *Glycobiology* **17**: 1150-5.

37. Jones J, Krag SS, Betenbaugh MJ (2005) Controlling N-linked glycan site occupancy. *Biochim. Biophys. Acta.* **1726**: 121-37.

CHAPTER 2

SEMI-SYNTHESIS OF IM7 GLYCOSYLATED-VARIANTS

Acknowledgements

The pTWIN(Im7$_{C29-G78}$) plasmid and the refolding protocol for the intein-Im7$_{C29-G78}$ construct were provided by Dr. Christian Hackenberger as described in:

Hackenberger CP, Chen MM, Imperiali B (2006) Expression of N-terminal Cys-protein fragments using an intein refolding strategy. *Bioorg. Med. Chem.*

The semi-synthetic strategy for the Im7 helix-I glyco-variants was designed by Dr. Christian Hackenberger as published in:

Hackenberger CP, Friel CT, Radford SE, Imperiali B (2005) Semisynthesis of a glycosylated Im7 analogue for protein folding studies. *J. Am. Chem. Soc.* **127**: 12882-9.

Although N-linked glycosylation is one of the most ubiquitous forms of protein modification (1), and occurs on approximately 50% of all eukaryotic proteins (2), comparatively little is known about how glycosylation affects the function and activity of proteins. In fact, less than 10% of the protein crystal structures currently deposited in the Protein Data Bank contain N-linked glycan chains (3). Among those structures the glycan is often disordered, with electron density for only the first two or three glycan residues attached. The major cause of this disparity can be attributed to the challenges encountered with obtaining homogeneously glycosylated protein samples. This difficulty arises because, unlike nucleic acids and amino acids, carbohydrates are not synthesized from a coding template. Instead, a group of glycosyltransferases and glycosidases of varied activity and specificity work together to form and modify the glycan, resulting in tremendous complexity and diversity in both glycan structure and site occupancy (1,4). Recombinant expression in other organisms or even changes in growth and cultivation conditions usually results in non-native or abhorrent glycosylation patterns (5,6). Chemistry can potentially provide a solution to this problem.

Various chemical and chemoenzymatic techniques have been developed to address the need for homogeneously glycosylated proteins (7-9). Each method has its inherent advantages and limitations with regards to site selection and protein yield.

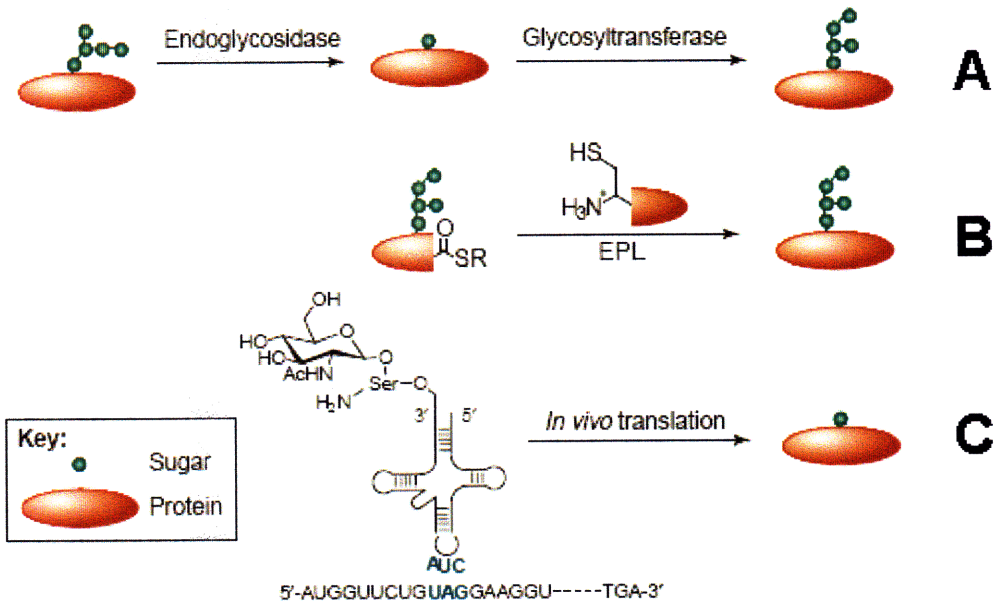


Figure 2-1. Comparison of three chemical approaches to glycoprotein synthesis: **A**) glycoprotein remodeling; **B**) semi-synthesis; and **C**) tRNA suppression.

Glycoprotein remodeling is one such method which begins with a heterogeneous sample of a glycoprotein and uses specific glycosidases to cleave off the heterogeneous portions of the glycan, usually leaving a common mono- or disaccharide (**Figure 2-1A**) (10,11). This truncated glycan can then be elaborated with specific glycosyltransferases to form the desired glycan. However, this method does not allow for site control and is therefore limited to only native glycosylation sites. It cannot guarantee homogeneity in glycan site occupancy, as often not all sites are 100% glycosylated in every protein. Furthermore, the glycosyltransferase reactions do not always proceed to completion, resulting in a mixture of glycoforms that may be difficult to separate chromatographically.

In vivo tRNA suppression is another method of glycoprotein synthesis which has been developed as a general strategy for incorporating unnatural amino acids into proteins through the use of the amber stop codon (TAG) and an unnatural amber suppressing amino acyl $tRNA_{(CUA)}$ (**Figure 2-1C**) (12,13). The main challenge of this method is the directed evolution of a tRNA synthetase which can selectively acylate an amber codon $tRNA_{(CUA)}$ with the glycosylated amino acid. Once that has been accomplished, this method allows for the site specific incorporation of the glycan at any site within the protein (14,15). Currently, this method has been used for the incorporation of β-GlcNAc-serine and β-GlcNAc-threonine into proteins expressed in *Escherichia coli* (16). However, this method has not yet been demonstrated for the synthesis of N-linked glycoproteins, most likely because the translation efficiency and yield of the glycoprotein is often limiting and can only be empirically determined.

Currently, the most robust and general approach to obtaining N-linked glycoproteins is through a semi-synthesis methodology using the expressed protein ligation (EPL) reaction (**Figure 2-2**) (17,18). This method involves the synthesis of a portion of the protein as a glycopeptide using solid phase peptide synthesis, which offers complete control of glycosylation site, as well as useful yields in the multi-milligram range. The remaining protein is recombinantly expressed. If the N-terminal peptide contains a C-terminal thioester and the C-terminal peptide contains an N-terminal cysteine, the two peptides can be joined by a chemical ligation reaction, which establishes a native amide bond between the two peptides.

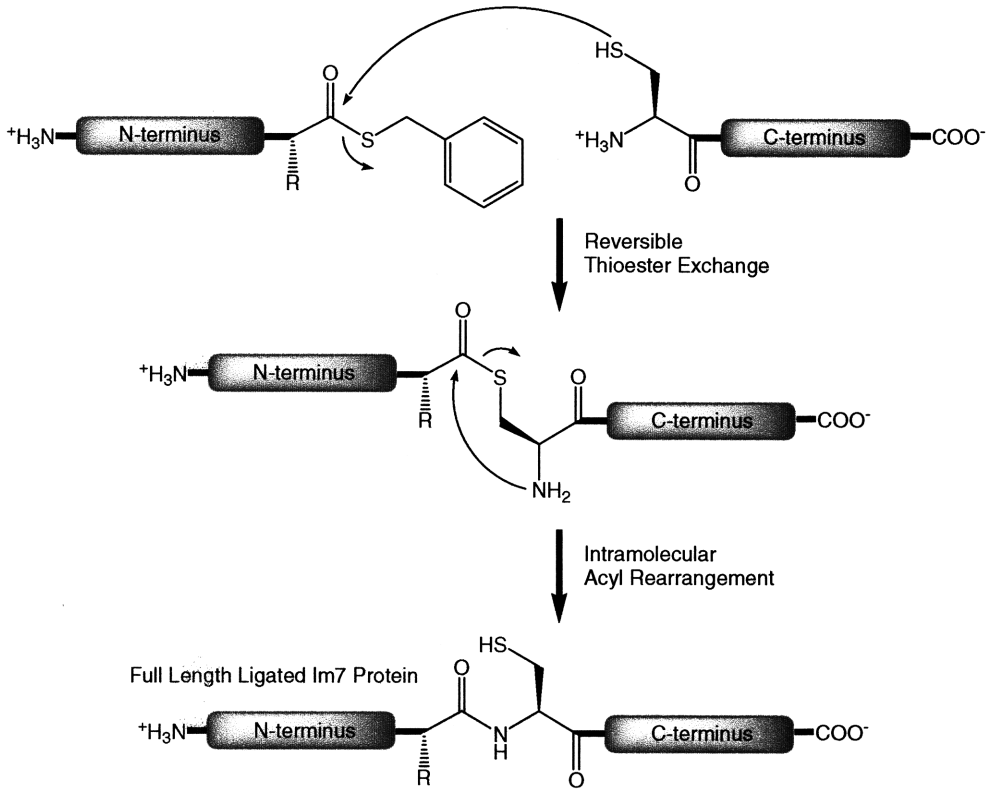


Figure 2-2. Native chemical ligation reaction mechanism (17).

The requirement for a cysteine residue at the site of ligation can be met by simply incorporating a cysteine point mutation at a site that does not impact protein function and structure, or alternatively a cysteine can be transiently incorporated to facilitate the EPL reaction and then chemically removed following the ligation, for example through Raney nickel desulfurization (19,20). The main advantages to semi-synthesis are the control of glycosylation site, the general applicability to all proteins, and the superior yield in comparison to other methods.

2-1. Semi-synthetic strategy of Im7 glyco-variants

For our current study we have developed a semi-synthetic strategy to obtain homogeneously glycosylated variants of a full length protein in order to study the effects of N-linked

glycosylation on protein folding kinetics. We have chosen a small model protein of four helices and 87 amino acids called immunity protein 7 (Im7). Im7 is a bacterial protein that functions as a binding partner for the bacterial toxin protein colicin E7 (**Figure 2-3**) (21,22). Both Im7 and colicin E7 are produced from the same plasmid, where Im7 protects the organism from the cytotoxic properties of E7 prior to its secretion. Im7 is an ideal model protein for our study because it has a well-characterized folding mechanism, and it lacks disulfide bonds, cis-proline amide linkages, and prosthetic groups that can complicate folding kinetics (23). Although Im7 does not contain any native glycosylation sites, the careful selection and introduction of glycans within various secondary structural motifs within this model protein can potentially reveal a great deal about the general effects of N-linked glycosylation. While the thermodynamic effects of N-linked glycosylation have been studied in peptide model systems (3,24,25), to our knowledge this will be the first systematic study aimed at a quantitative characterization of the folding kinetics of a full length homogeneously glycosylated protein.

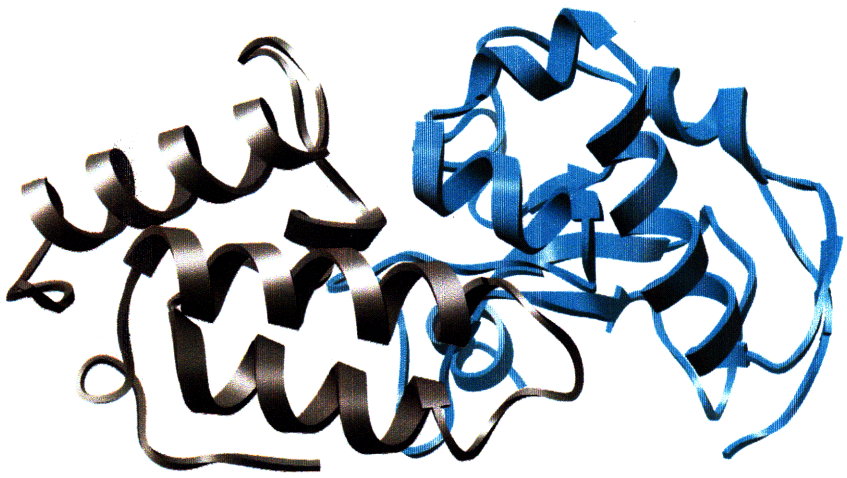


Figure 2-3. Co-crystal structures of immunity protein Im7 (dark) and collicin E7 (blue) in ribbon representation from RCSB Protein Data Bank (pdb code 7CEI).

In collaboration with Professor Sheena Radford's lab at the University of Leeds, we selected seven sites of interest within the first and last 28 amino acids (**Figure 2-4**). We limited the sites to solvent exposed residues in order to minimize any obvious steric clashes that would prevent the protein from folding to its native structure. Among the seven selected glycosylation sites, two are located within the middle of a helix, two are located near the end of a helix, and three are located within loop regions of different topology. This range of glycosylation sites will allow us to quantitatively determine whether the effects of N-linked glycosylation are general or specific to the local secondary structure.

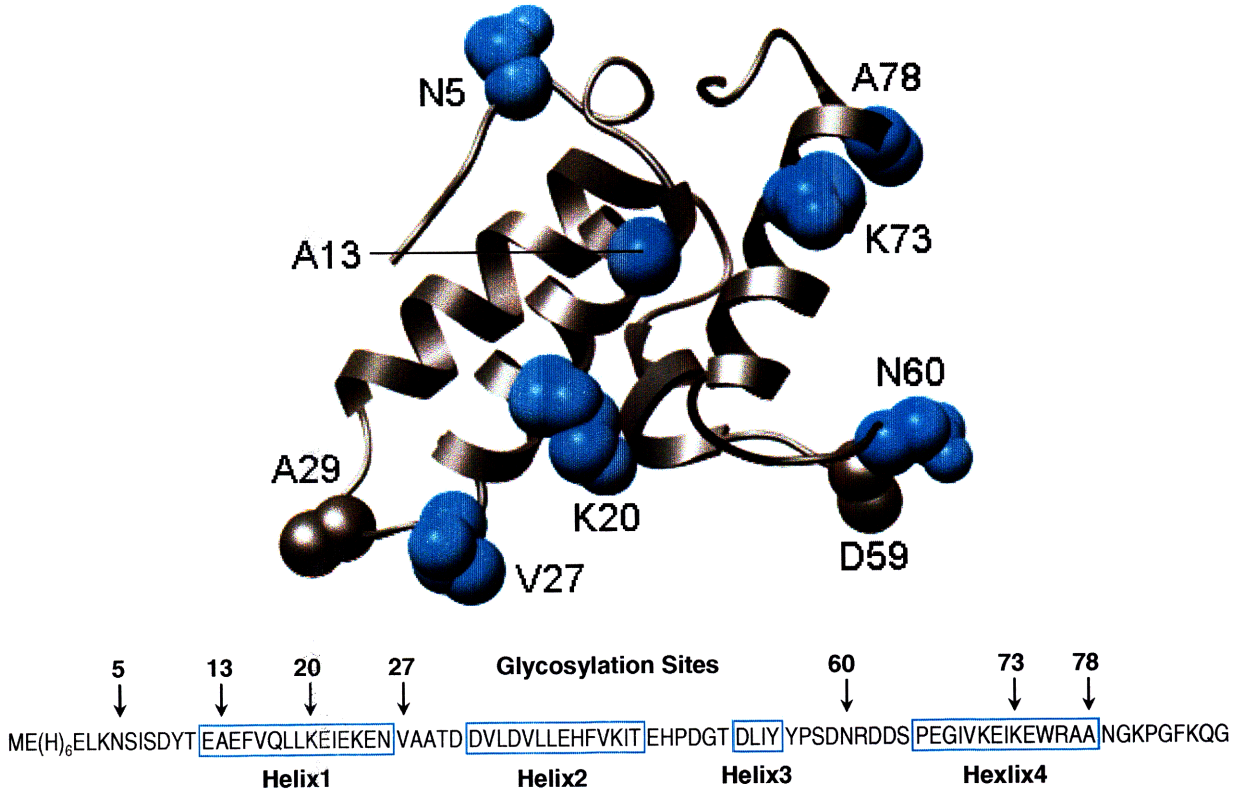


Figure 2-4. Ribbon diagram of Im7 (pdb code 1AYI) with ligation (dark) and glycosylation (blue) sites highlighted with spacefill side-chain representation. Amino acid sequence is listed below with the four helices marked by boxes.

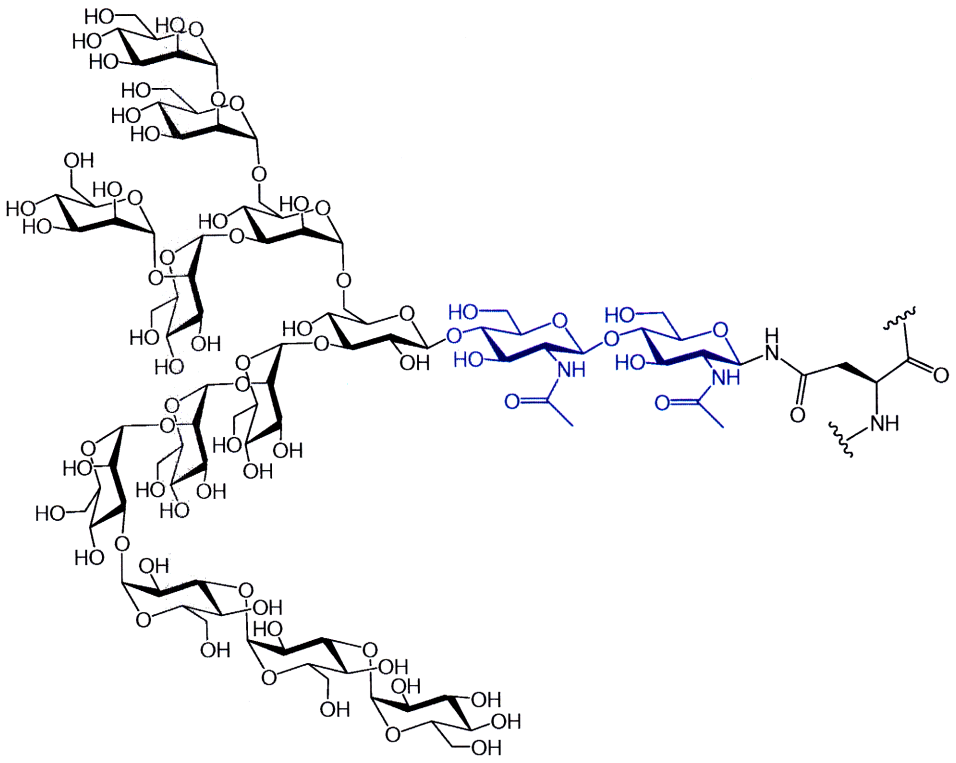


Figure 2-5. The tetradecasaccharide ($Glc_2Man_9GlcNAc_2$) from *Saccharomyces cerevisiae* attached to the asparagine side chain of the acceptor protein. The chitobiose disaccharide, found in all eukaryotic N-linked glycans is highlighted in blue. Wavey lines indicate the extension of the polypeptide chain. Structure drawing adapted from (24).

In eukaryotic N-linked glycosylation, a tetradecasaccharide ($Glc_2Man_9GlcNAc_2$, **Figure 2-5**) is initially transferred to proteins, followed by post-translational processing in the ER and the Golgi apparatus where significant changes in the glycan structure accumulate (1,4). Despite these modifications, the disaccharide that is attached to the asparagine, GlcNAc-β-1,4-GlcNAc-β, often referred to as chitobiose, remains unchanged. Using various glycopeptides, previous studies from our lab have demonstrated significant structural effects on the peptide backbone originate from the chitobiose moiety, while the distal sugars only serve to amplify the initial effect (26,27). Interestingly, this disaccharide alone is capable of inducing the formation of a compact beta-turn (26,28). Since the chitobiose disaccharide is common to all N-linked

glycoproteins and since it displays a significant effect on the local peptide structure, we used this truncated disaccharide as the model glycan in the Im7 glycoprotein assembly.

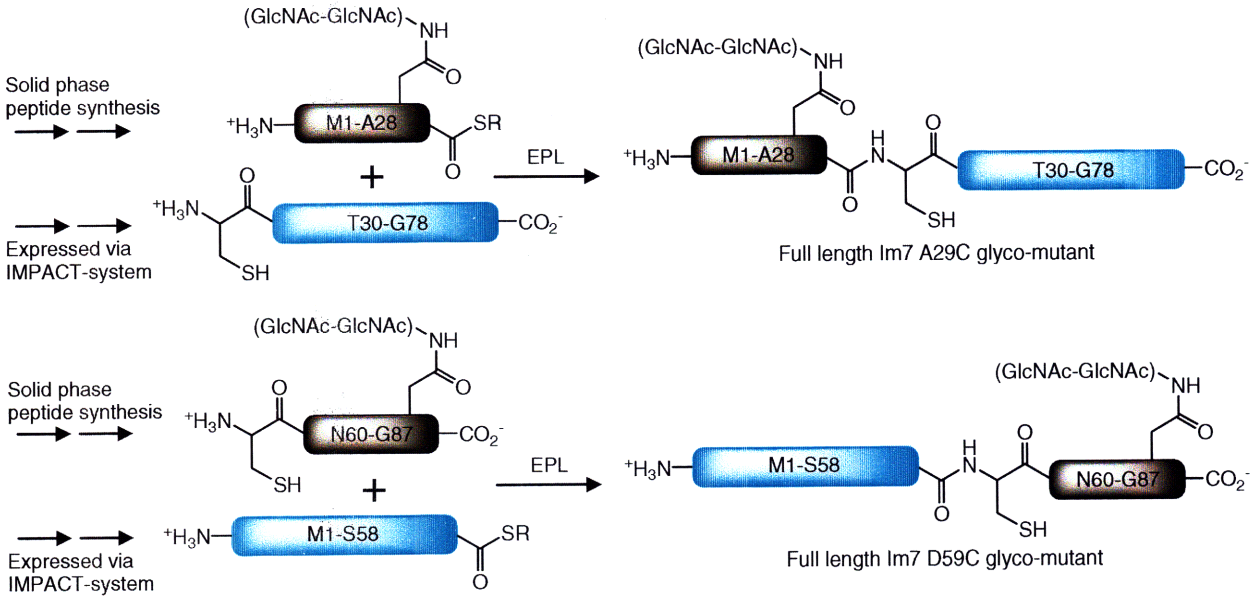


Figure 2-6. Im7 glyco-variants semi-synthetic strategy. Each variant consists of a glycopeptides made by solid phase peptide synthesis and a larger peptide made by expression. EPL = expressed protein ligation.

The semi-synthetic strategy we employ differs depending on whether the glycosylation site is near the N- or C-terminus of the protein (**Figure 2-6**). In both cases, the glycopeptide was synthesized by solid phase peptide synthesis, while the two larger peptides were both recombinantly expressed with intein fusion tags from the commercially available IMPACT-system (29) in order to incorporate the necessary N-terminal cysteine or the C-terminal thioester. A cysteine was introduced at the A29C position to accommodate the N-terminal glycoproteins, and another at the D59C position to accommodate the C-terminal glycoproteins. Both cysteines were chosen to occupy sites that allow their side chains to be solvent exposed and were not expected to form inter-residue contacts in the native state. Since they were not expected to

perturb the structure or folding of the overall protein, the cysteine residues were not removed after ligation. For each corresponding Im7 glyco-variant, a non-glyco pseudo-wildtype was also prepared, which contained both the cysteine mutation as well as the non-glycosylated asparagine. Therefore, the direct effects of the glycan can be determined by comparing the glyco-variant with the pseudo-wildtype control.

2-2. Synthesis of the protected Fmoc-Asn(chitobiose-TBDMS$_5$)-OH building block

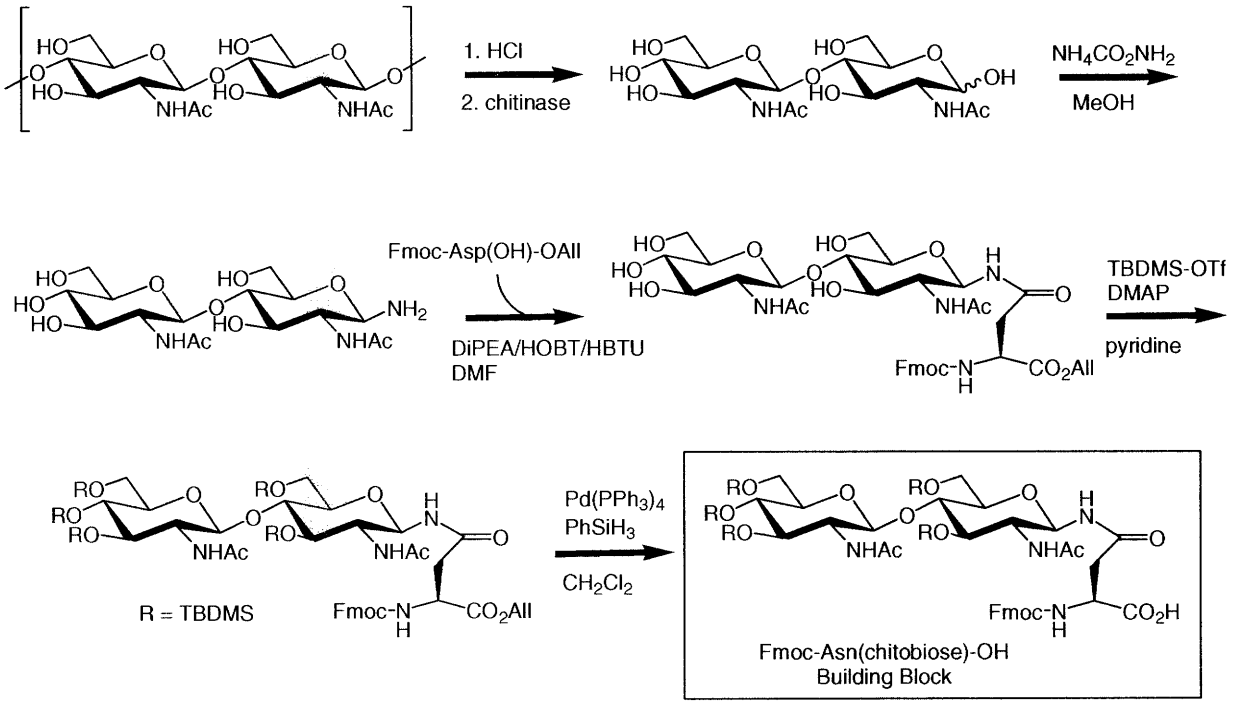


Figure 2-7. Synthesis of Fmoc-Asn(chitobiose-TBDMS$_5$)-OH building block.

Synthesis of the protected building block was carried out according to a previously developed protocol from our lab (**Figure 2-7**) (30). Briefly, the chitobiose disaccharide was produced by enzymatic degradation of chitin using a commercially available chitinase. The chitobiose was

then aminated at the anomeric position by the Likhosherstov method using ammonium carbamate in methanol (31). The resulting glycosylamine was precipitated and immediately coupled to the acid side chain of the Fmoc-Asp-OAll amino acid. Each hydroxyl group of the chitobiose was silyl protected using TBDMS-OTf (*tert*-butyldimethylsilyltrifluoromethane sulfonate). Final deprotection of the allyl group resulted in the desired building block product in a form compatible with standard Fmoc-based solid-phase peptide synthesis. The key improvement of this synthesis over previous methods was the Likhosherstov amination (31), which is stereospecific for the beta-glycosylamine, since the alpha-glycosylamine is not biologically relevant (30).

2-3. Synthesis of the $Im7_{C59-G87}$ glycopeptides and the $Im7_{M1-A28}$ glycopeptide thioesters

The $Im7_{M1-A28}$ and $Im7_{C59-G87}$ glycopeptides (**Figure 2-8**) were synthesized by solid phase peptide synthesis (SPPS) using standard Fmoc-based conditions with 2-(1*H*-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate and hydroxybenzotriazole (HBTU and HOBt) as the primary coupling reagents. However, 7-azabenzotriazol-1-yloxy-tris(pyrrolidino)phosphonium hexafluorophosphate (PyAOP) was used for the coupling of the Asn-chitobiose building block in order to accommodate the steric demands imposed by this bulky amino acid (32). The highly acid-labile carboxy-trityl TGT resin was used as the solid support. In the synthesis of $Im7_{M1-A28}$ glycopeptide thioester, this resin selection played a critical role by allowing the completed peptide to be cleaved from the resin under mild acidic conditions with all the side chain protecting groups of the peptide intact. The free C-terminal acid can then be converted into a thioester through chemical synthesis (33).

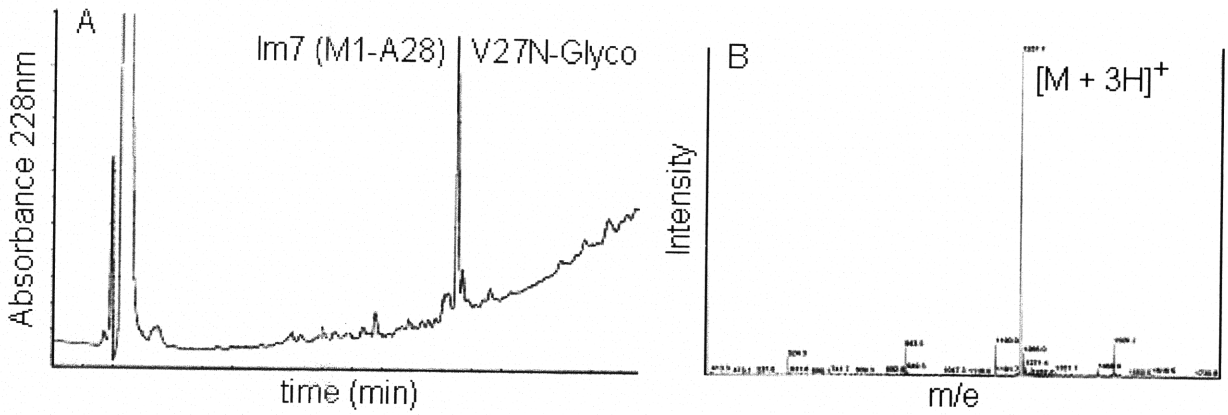


Figure 2-8. $Im7_{M1-A28}$ V27N-glycopeptide: **A**) Reverse phase preparative C_{18} HPLC profile at 228 nm and **B**) ESI mass spectra.

2-4. Expression of the $Im7_{C29-G78}$ peptide.

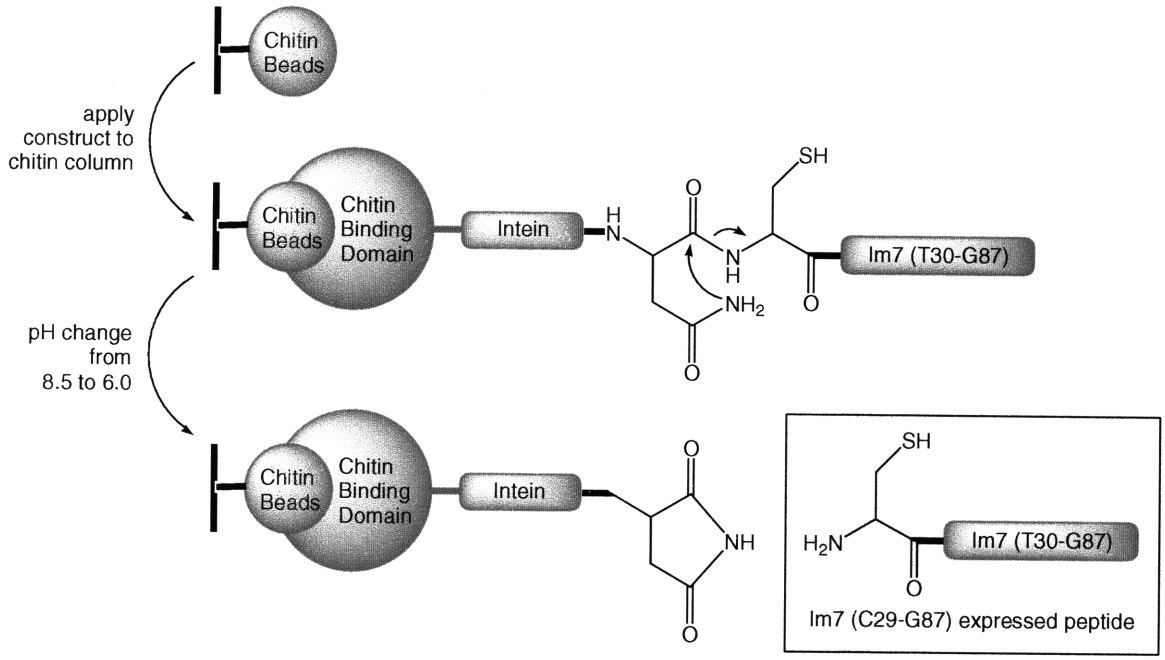


Figure 2-9. Intein cleavage mechanism and the production of the $Im7_{C29-G87}$ peptide.

Expression of peptides containing an amino acid other than methionine at the N-terminus can be achieved using a cleavable recombinant expression construct. In this case we used the

commercially available pTWIN1 plasmid from New England Biolabs IMPACT-system, which produced a construct consisting of a chitin binding domain for affinity purification, an engineered pH-controlled intein, followed by the desired protein (**Figure 2-9**). Purification of this construct was carried out at pH 8.5 (34-36). Once the construct had been bound to the chitin resin and thoroughly washed, it was incubated overnight at pH 6, which caused the intein to self-cleave, releasing the product with any desired N-terminal amino acid -- in this case a cysteine (**Figure 2-10**). The eluted $Im7_{C29\text{-}G78}$ peptide was further purified and desalted by reverse phase HPLC (**Figure 2-11**), aliquoted, and lyophilized for storage. Not only does this construct allow for the expression of a protein with a non-methionine N-terminal amino acid, it was also believed to help solubilize truncated proteins that have exposed hydrophobic core residues and reduce their aggregation (36).

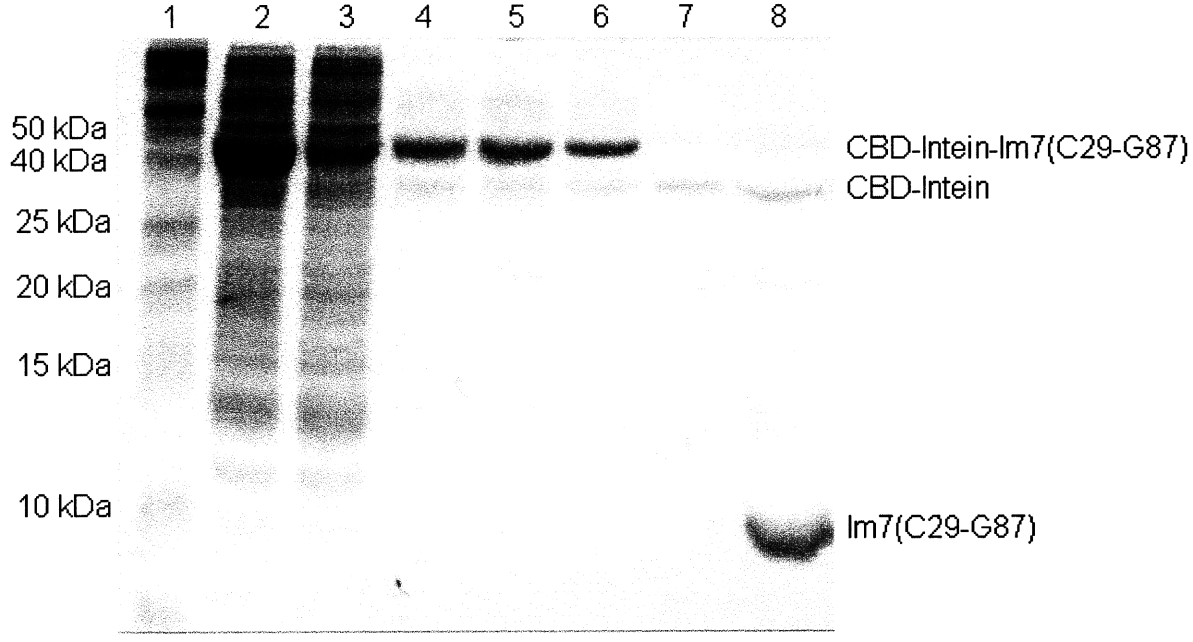


Figure 2-10. SDS-PAGE gel of $Im7_{C29\text{-}G87}$ purification. Lane 1: gel ladder; Lane 2: cell lysate after dialysis refolding; Lane 3: flow through chitin column; Lane 4-6: wash with buffer pH = 8.5; Lane 7-8: elution with buffer pH = 6 after 12 hour incubation.

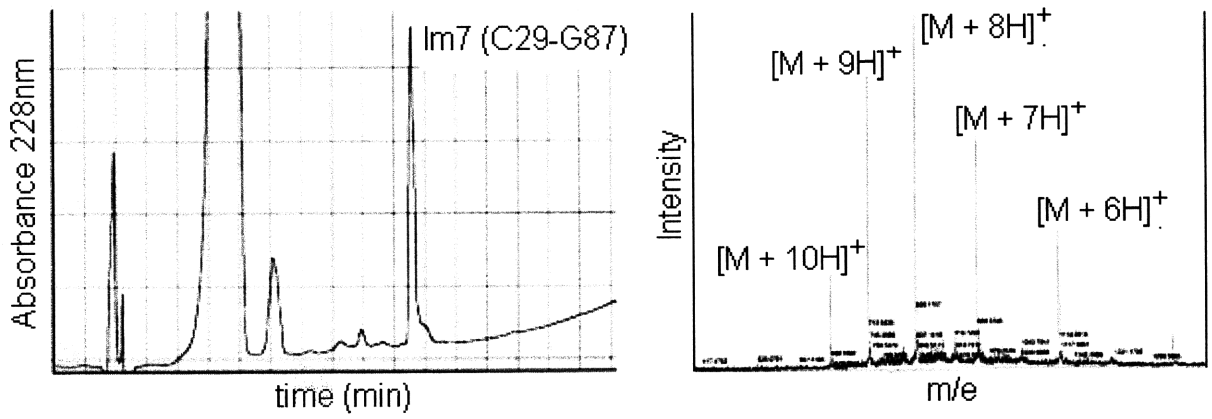


Figure 2-11. $Im7_{C29\text{-}G87}$ peptide: **A**) Reverse phase preparative C_{18} HPLC profile at 228 nm and **B**) ESI mass spectra.

2-4. Expression of the $Im7_{M1\text{-}S58}$ peptide thioester.

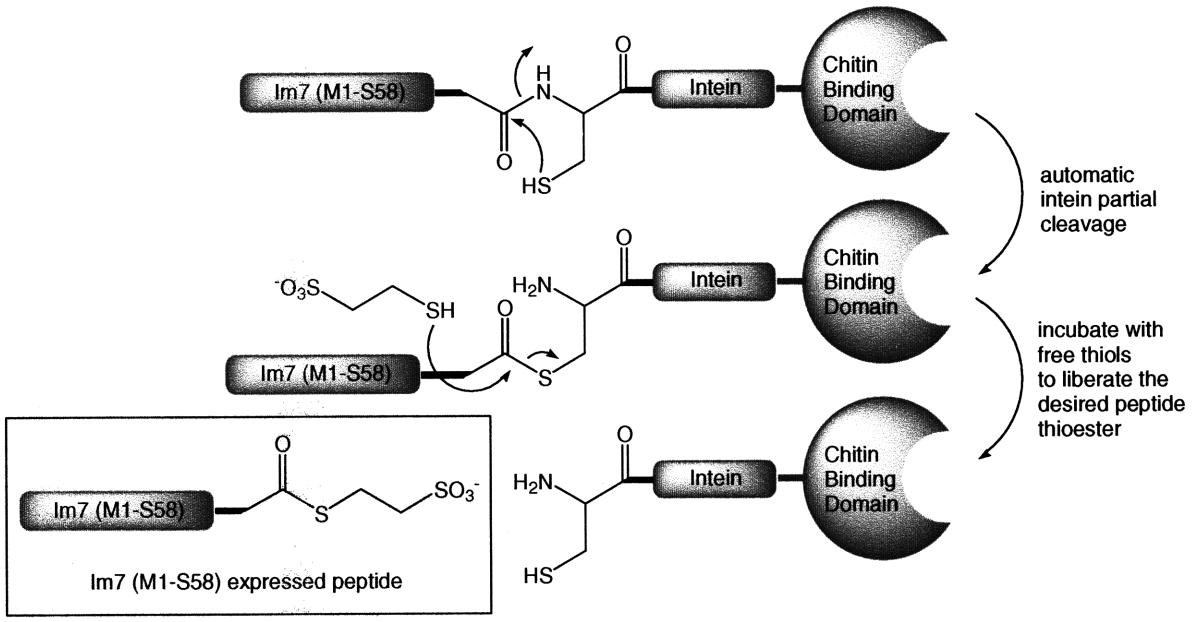


Figure 2-12. Intein cleavage mechanism and the production of the $Im7_{M1\text{-}S58}$ peptide.

Expression of the $Im7_{M1\text{-}S58}$ peptide thioester was also carried out using intein technology from the pTWIN1 plasmid from New England Biolabs. In this case a construct was produced,

consisting of the desired peptide, an engineered partially functional intein, followed by the chitin binding domain for affinity purification (**Figure 2-12**) (34,35). The intein in this construct contained a cysteine to alanine mutation which caused it to stop the cleavage mechanism after the initial rearrangement. This left a thioester linkage between the desired peptide and the intein, which allowed the desired peptide to be liberated as a free thioester through incubation with a high solution concentration of free thiol (**Figure 2-13**) (35).

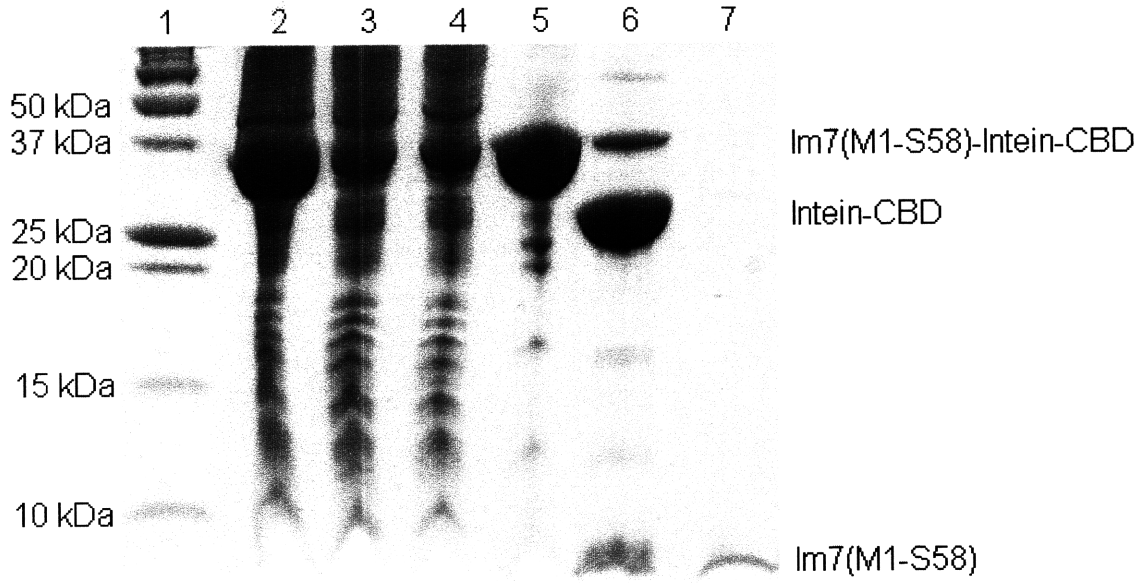


Figure 2-13. SDS-PAGE gel of $Im7_{M1-S58}$ purification. Lane 1: gel ladder; Lane 2: cell lysate; Lane 3: flow through Ni-NTA column; Lane 4: Ni-NTA column wash; Lane 5: Ni-NTA elution; Lane 6: MESNA 8 hour incubation; Lane 7: flow through chitin column.

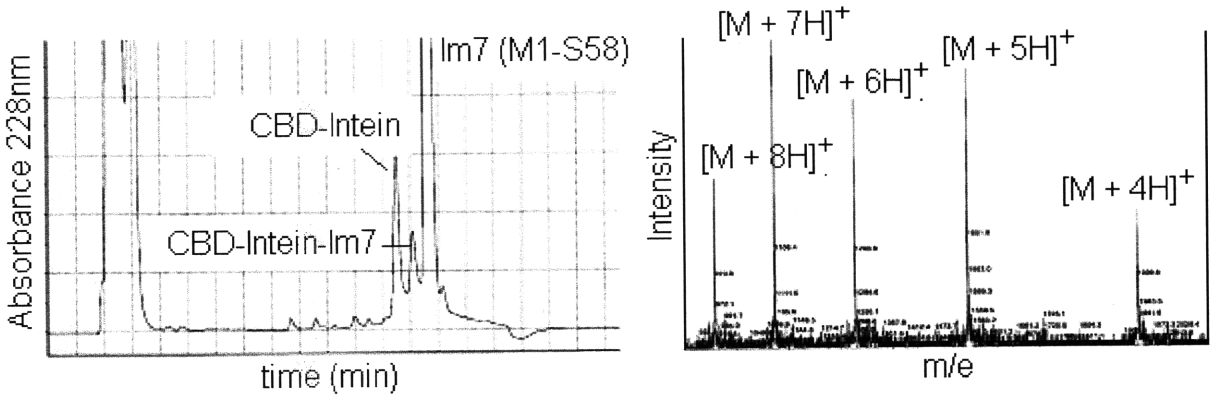


Figure 2-14. $Im7_{M1-S58}$ peptide thioester: **A**) Reverse phase preparative C_{18} HPLC profile at 228 nm and **B**) corresponding ESI mass spectra.

2-6. Native chemical ligation

Ligation of the synthetic peptide with the corresponding expressed peptide was carried out under identical conditions for both the N-terminal glyco-variants as well as the C-terminal ones. Both ligation sites were chosen with efficiency in mind, avoiding residues with large or branched side chains in order to minimize steric hindrance during ligation. As a result, all ligations were successfully carried out under native conditions with approximately 80% yield relative to the limiting glycopeptides (32). The full length Im7 product can be efficiently purified from the peptide starting materials by preparative C_{18} reverse phase HPLC (**Figure 2-15**). Although the conditions of the HPLC no doubt caused the protein to denature, the robust quality of Im7 allowed it to successfully refold into its native structure when returned to a neutral aqueous buffer.

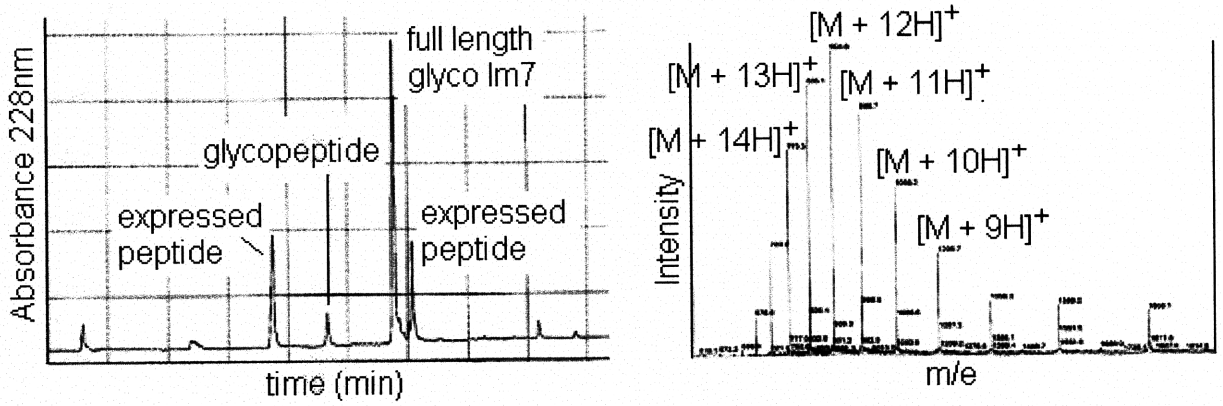


Figure 2-15. Full length ligated $Im7_{A29C/V27NGlyco}$ protein: **A**) Reverse phase preparative C_{18} HPLC profile at 228 nm and **B**) corresponding ESI mass spectra. It is not clear why the expressed peptide has two elution peaks, although one possibility is that one is a monomer and the other is a disulfide dimer.

2-7. Expression of the non-glyco pseudo-wildtype Im7 controls

For each glyco-variant, a corresponding non-glyco pseudo-wildtype Im7 control was produced through bacterial expression. Using site direct mutagenesis, each pseudo-wildtype included the ligation cysteine mutation at either position 29 or 59, as well as the asparagine mutation if the glycosylation site is at a site that was not originally an asparagine. Therefore, each glyco-variant of Im7 differed from its pseudo-wildtype only by the presence of the chitobiose disaccharide. When directly compared, each Im7 pair should reveal the effect of the glycan alone.

2-8. Characterization of Im7 variants

In total, approximately 6-12 mg of each Im7 glyco-variant was produced using our semi-synthetic strategy, along with more than 20 mg of each corresponding pseudo-wildtypes using site directed mutagenesis. The purity and structural attributes of each variant were thoroughly assessed prior to folding studies. Electrospray-ionization mass spectrometry established the

mass of each variant to be consistent with expected values. Each glyco-variant differed from the non-glyco-variant by precisely the mass of the glycan moiety. Furthermore, no major extraneous peaks were observed in the HPLC or mass spectrometry analysis. SDS-PAGE electrophoresis revealed each sample to be greater than 95% pure. Close inspection of each lane of the gel loaded with the glyco-variant did not indicate the presence of the ligation starting material or the deglycosylated protein.

Circular dichroism spectroscopy was used to compare the secondary structures of each protein (**Figure 2-16**). Since Im7 contained four helices, the signal at 208 and 222 nm is especially indicative of its fold. Within error, we did not find the glycan nor the mutations incorporated in our new constructs to significantly alter the helical content of Im7. Although this does not offer information on the tertiary structure of Im7, it is consistent with our expectation that each variant fold to the same native state (32,37).

To further investigate the structure of the folded state of each variant, we compared the fluorescence of W75 within Im7 under native and denaturing conditions of 8 M urea (**Figure 2-17**). Residue W75 is highly fluorescent in the unfolded state. However, in the native state, it is placed in close proximity to H47 which quenches its fluorescence (32,37). The same degree of fluorescence quenching observed for each Im7 further suggested that each protein is folded to the same native state since this quenching phenomenon is highly dependent on the relative positions of W75 and H47.

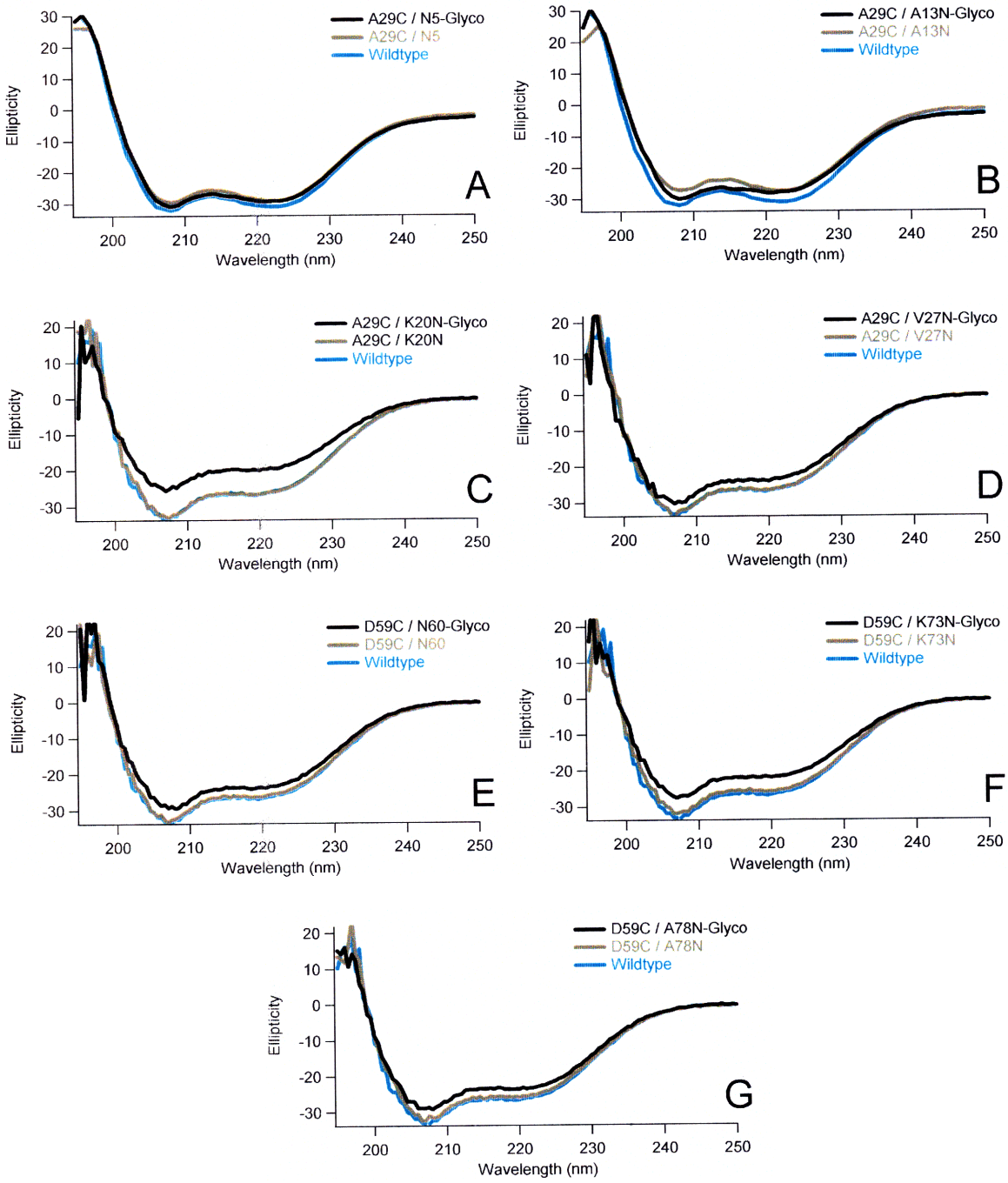


Figure 2-16. Circular dichroism spectra of Im7 variants.

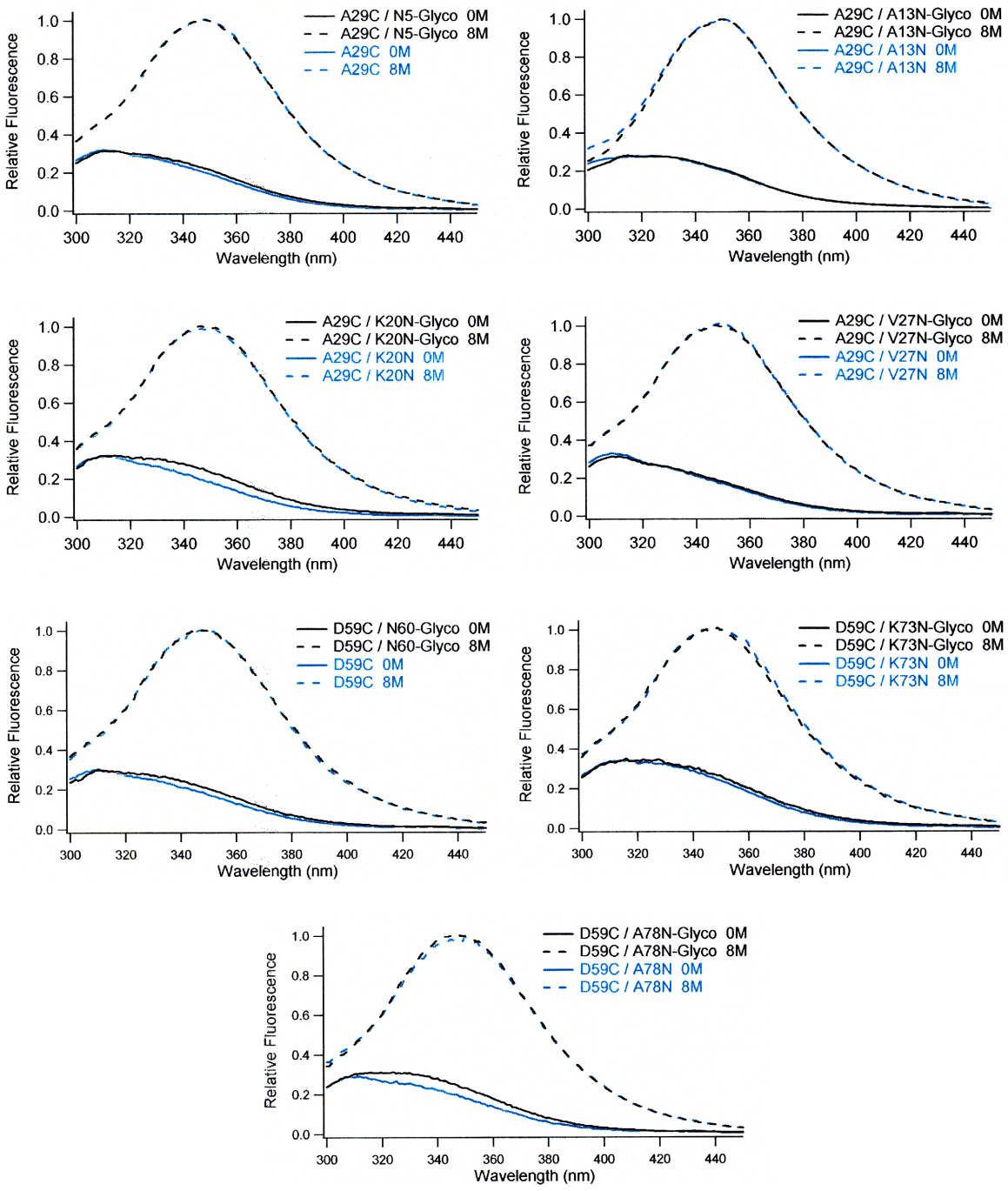


Figure 2-17. Fluorescence emission spectra of Im7 variants in 0 and 8 M urea at λ_{max} = 280 nm.

Conclusion

In conclusion, we have developed a general and robust strategy for the semi-synthesis of homogeneously glycosylated Im7 variants. Our strategy involved the ligation of a synthetic glycopeptide with a larger recombinantly expressed peptide. This approach allowed us to incorporate a biologically relevant disaccharide anywhere within the first or last 28 amino acids of our protein, and prepare the proteins in multi-milligram quantities of greater than 95% purity. We applied this strategy to access seven glyco-Im7 variants, placing the glycan at seven carefully selected locations representing a survey of different secondary structures from helices to loops. Circular dichroism and fluorescence spectroscopy was used to verify that the mutations and glycans have not significantly changed the native folded state of Im7. Most importantly, we now have access to enough protein to carry out stopped-flow fluorescence spectrometry experiments to quantitatively measure the effects of the glycan on the kinetics of protein folding.

Our semi-synthetic strategy represents one of the most general methods of glycoprotein synthesis currently reported. It allowed for the production of glycoprotein in sufficient quantities and to a degree of purity and homogeneity to allow the use of various biophysical techniques for their detailed characterization. Successful application of this strategy for the synthesis of other glycoproteins may help to address our current limitations in the study of protein glycosylation.

Experimental

Synthesis of Fmoc-Asn(Chitobiose-TBDMS$_5$)-OH.

Synthesis of the Fmoc-Asn(Chitobiose-TBDMS$_5$)-OH building block was carried out following literature protocol (30) starting from commercially available crab shell chitin. The critical step in the synthesis involves the drying of the isolated crude amino chitobiose prior to coupling to the Fmoc-Asp(OH)-OAll amino acid. Excessive drying causes degradation, while inadequate drying leaves residual methanol which interferes with the subsequent coupling reaction.

Sequence, numbering, and structure of the Im7 construct.

MEHHHHHH E^2LKNSISDY TEAEFVQLLK EIEKENVAAT DDVLDVLLEH
FVKITEHPDG TDLIYYPSDN RDDSPEGIVK EIKEWRAANG
KPGFKQG

The residue numbers listed in this thesis are those of the untagged protein, where (E^2) corresponds to the second residue of the protein (38). The crystal structure of Im7 protein can be found at the RCSB Protein Data Bank code 1AYI (22).

Synthesis and purification of the Im7$_{C59\text{-}G87}$ glycopeptides.

The Im7$_{C59\text{-}G87}$ glycopeptides were synthesized on Fmoc-Gly-NovaSyn TGT resin (NovaBiochem) starting on an ABI 431A automated peptide synthesizer (Applied Biosystems) using standard Fmoc (9-fluoenylmethoxycarbonyl) coupling procedures. Fmoc deprotection was carried out using 20% 4-methylpiperidine in NMP (1-methyl-2-pyrrolidinone, Aldrich). Double coupling for each amino acid was carried out using 4 eq. Fmoc-amino acid (GenScript or NovaBioChem) per eq. resin, 4 eq. HOBt (N-Hydroxybenzotriazole, GenScript), and 4 eq. HBTU (2-(1HBenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate, GenScript)

as activating agents, with DIPEA (diisopropylethylamine, Aldrich) in NMP, for one hour at room temperature. Acetic anhydride capping was carried out to terminate unreacted free amines after each amino acid.

Coupling of the Fmoc-Asn(Chitobiose-$TBDMS_5$)-OH building block, as well as the 2 subsequent amino acids, were carried out by hand using 4 eq. of the Fmoc-amino acid, and 4 eq. PyAOP (7-azabenzotriazol-1-yloxy-tris(pyrrolidino)phosphonium hexafluorophosphate, GenScript) as the coupling agent, with DIPEA in DMF (dimethylformamide, Aldrich). TNBS (2,4,6-trinitrobenzenesulfonic acid) was used to check for the presence of a free amino terminus for completion of each reaction.

Cleavage from the resin and global deprotection was carried out using cleavage cocktail K [82% TFA (trifluoroacetic acid), 3% EDT (1,2 ethanedithiol), 5% thioanisol, 5% phenol, and 5% water], shaking for 2 hours at room temperature. The desired glycopeptide was precipitated in cold diethyl ether, purified by preparative reverse phase C_{18} HPLC, confirmed by electrospray ionization mass spectrometry (ESI-MS) to be within 1 Da of expected mass, checked by Coomassie staining of SDS-PAGE to be ≥95% pure, and lyophilized for storage.

Synthesis and purification of the $Im7_{M1-A28}$ glycopeptide thioester.

The $Im7_{M1-A28}$ glycopeptides were synthesized on Fmoc-Ala-NovaSyn TGT resin (NovaBiochem) using the same protocol as described for the $Im7_{C59-G87}$ peptide above. When the glycopeptide was completed, it was cleaved from the resin using cleavage cocktail A [0.5% TFA, 99.5% dichloromethane], shaking for 2 hours at room temperature, resulting in a peptide

with a free carboxyl terminus and fully intact side chain protecting groups. This peptide was dried under reduced pressure, washed with diethyl ether, redissolved in anhydrous DMF, and treated with 10 eq. of benzyl mercaptan (Aldrich), 3 eq. of PyBOP (benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate, GenScript), followed by 3 eq. of DIPEA. The mixture was stirred for 8 hours under N_2. The DMF solvent was removed by rotoary evaporation under high vacuum, and the resulting yellow oil is treated for 2 hours at room temperature with cleavage cocktail K for global deprotection. The desired glycopeptide thioester were precipitated in cold diethyl ether, purified by preparative reverse phase C_{18} HPLC, confirmed by ESI-MS to be within 1 Da of expected mass, check by Coomassie staining of SDS-PAGE to be ≥95% pure, and lyophilized for storage.

Mutagenesis of Im7 non-glyco pseudo-wildtype variants.

A pTrc(Im7)-99A vector was provided by Professor Sheena Radford from the University of Leeds, encoding the native Im7 protein with an N-terminal hexa-histidine tag described elsewhere (38). Introduction of point mutations in Im7 was performed using the Quikchange mutagenesis kit (Stratagene). All variants were sequenced to ensure that the gene contained the desired change.

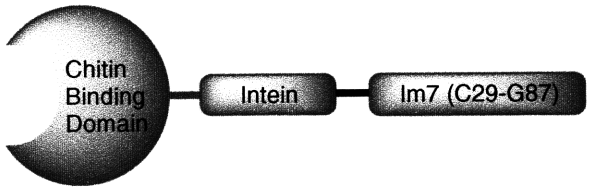
Expression and purification of Im7 non-glyco pseudo-wildtype variants.

Starting from a 5 mL overnight culture, BL21(DE3) cells (Stratagene) expressing Im7 variants were grown at 37 °C in 1 L of LB broth containing carbenicillin antibiotic to an OD_{600} of 0.6-0.8. At that point protein production was induced by the addition of 1 mM IPTG (isopropyl-β-D-thiogalactopyranoside). After 3 hours, the cells were harvested by centrifugation (7,500 x *g*) for

30 min, washed once with a 0.9% NaCl solution, centrifuged again (7,500 x *g*) for 30 min, and the cell pellet was frozen at -80 °C until needed.

All purifications steps were performed at 4 °C. Cell pellets containing the Im7 variants were thawed and resuspended in 5% of the original culture volume in buffer L [50 mM Tris-acetate, 5 mM imidazole, pH 8]. The cells were lysed by sonication, followed by centrifugation (142,400 x g) for 1 h to remove cellular debris and membrane proteins. The supernatant was slowly applied to a column containing Ni-NTA agarose equilibrated with buffer L. After washing with 5 column volumes of buffer L, the purified protein was eluted with buffer E [50 mM Tris-acetate, 250 mM imidazole, pH 8]. Fractions containing a significant amount of desired protein were combined and purified by preparative reverse phase C_{18} HPLC, confirmed by ESI-MS to be within 1 Da of expected mass, checked by Coomassie staining of SDS-PAGE to be ≥95% pure, and lyophilized for storage.

Cloning of Intein Im7$_{C29\text{-}G87}$ construct.



Starting from the pTrc(Im7$_{A29C}$)-99A vector, the Im7$_{C29\text{-}G87}$ peptide was amplified using primers which engineered 5' Sap1 and 3' BamH1 restriction sites and the Vent polymerase (New England Biolabs) under standard conditions described by the manufacturer for 30 cycles. Both the amplicons and the IMPACT pTWIN1 vector was doubly digested with Sap1 and BamH1 restriction enzymes (New England Biolabs), fractionalized by agarose gel electrophoresis, and

purified with the QIAquick gel extraction kit (Qiagen). Ligations were conducted with the T4 DNA ligase kit (Promega) for 1.5 hours at 16 °C. The ligation product was transformed into a competent DH5α strain of *E. coli* (Invitrogen), spread on plates containing LB media and carbanicillin. The desired pTWIN1(Im7$_{C29\text{-}G87}$) plasmid from colonies formed on the plates were isolated, verified by sequencing at the MIT CCR HHMI Biopolymers Lab, and transformed into a competent BL21(DE3) strain of *E. coli* for overexpression.

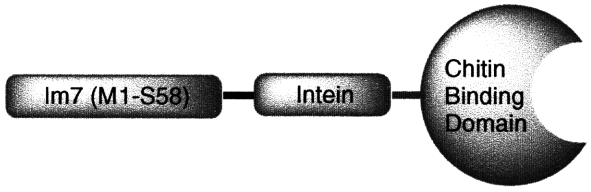
Expression and purification of Im7$_{C29\text{-}G87}$ peptide.

Starting from a 5 mL overnight culture, BL21(DE3) cells (Stratagene) expressing pTWIN1(Im7$_{C29\text{-}G87}$) were grown at 37 °C in LB broth containing carbenicillin antibiotic to an OD_{600} of 0.6-0.8. At that point protein production was induced by the addition of 1 mM IPTG (isopropyl-β-D-thiogalactopyranoside). Expression at this temperature leads to the overexpressed construct to form inclusion bodies. After 3 hours, the cells were harvested by centrifugation (7,500 x *g*) for 30 min, washed once with a 0.9% NaCl solution, centrifuged again (7,500 x *g*) for 30 min, and the cell pellet was frozen at -80 °C until needed.

All purifications steps were performed at 4 °C. Cell pellets containing the Im7$_{C29\text{-}G87}$ construct were thawed and resuspended in 5% of the original culture volume in buffer M [20 mM Tris-HCl, 500 mM NaCl, pH 8.5]. The cells were lysed by sonication, followed by centrifugation (142,400 x g) for 1 hour to precipitate the cellular debris and inclusion bodies containing the desired construct. The isolated pellet was homogenized into 10% of the original culture volume in buffer G [20 mM Tris-HCl, 500 mM NaCl, 7 M guanidinium chloride, pH 8] and left stirring for 1 hour. The cellular debris was removed from the dissolved protein by centrifugation

(142,400 x *g*) for 1 hour. The supernatant was dialyzed into refolding buffers R-1/2/3/4/5/6 [20 mM Tris-HCl, 500 mM NaCl, 8/6/4/2/0/0 M urea, 1 mM DTT (dithiothreitol), pH 8.5] which contained sequentially dilute concentrations of urea, for 12 hours in each buffer. The dialysis product was then added to a column of chitin resin (New England Biolabs), of 1% the original culture volume and prewashed with buffer R-6, at a flow rate of 0.5 mL/min. The column was washed with 15 column volumes of buffer R-6 and then incubated in buffer C [20 mM Tris-HCl, 1 mM DTT, pH 6.0] at room temperature for 16 hours to induce intein cleavage. The column is then washed with buffer C until all protein products have been eluted. The isolated $Im7_{C29\text{-}G87}$ peptide is then further purified by preparative reverse phase C_{18} HPLC, confirmed by ESI-MS to be within 1 Da of expected mass, checked by Coomassie staining of SDS-PAGE to be ≥95% pure, and lyophilized for storage. A 1 liter expression typically yields approximately 4-5 mg of desired product.

Cloning of $Im7_{M1\text{-}S58}$ Intein construct.



Starting from the pTrc(Im7)-99A vector, the $Im7_{M1\text{-}S58}$ peptide was amplified using primers which engineered 5' Nde1 and 3' Sap1 restriction sites and the pFu Turbo polymerase (Stratagene) under standard conditions described by the manufacturer. Both the amplicons and the IMPACT pTWIN1 vector were doubly digested with Nde1 and Sap1 restriction enzymes (New England Biolabs), fractionalized by agarose gel electrophoresis, and purified with the QIAquick gel extraction kit (Qiagen). Ligations were conducted with the T4 DNA ligase kit

(Promega) for 15 min at room temperature. The ligation product was transformed into a competent DH5α strain of *E. coli* (Invitrogen), spread on plates containing LB media and carbanicillin. The desired pTWIN1($Im7_{M1-S58}$) plasmid from colonies formed on the plates were isolated, verified by sequencing at the MIT CCR HHMI Biopolymers Lab, and transformed into a competent BL21(DE3) strain of *E. coli* for overexpression.

Expression and purification of $Im7_{M1-S58}$ peptide thioester.

Starting from a 5 mL overnight culture, BL21(DE3) cells (Stratagene) expressing pTWIN1($Im7_{M1-S58}$) were grown at 37 °C in LB broth containing carbenicillin antibiotic to an OD_{600} of 0.6-0.8. At that point the temperature was lowered to 16 °C and protein production was induced by the addition of 1 mM IPTG for 16 hours. The cells were harvested by centrifugation (7,500 x *g*) for 30 min, washed once with a 0.9% NaCl solution, centrifuged again (7,500 x *g*) for 30 min, and the cell pellet was frozen at -80°C until needed.

All purifications steps were performed at 4 °C. Cell pellets containing the $Im7_{M1-S58}$ construct were thawed and resuspended in 2.5% of the original culture volume in buffer N [20 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), pH 8]. The cells were lysed by sonication, followed by centrifugation (142,400 x g) for 1 hour to remove cellular debris and membrane proteins. The supernatant was slowly applied to a column containing Ni-NTA agarose equilibrated with buffer L. After washing with 4 column volumes of buffer W [20 mM HEPES, pH 6], the purified construct was eluted with 25 mL of buffer D [20 mM HEPES, 200 mM imidazole, pH 6]. To the elution was added 2 g MESNA (sodium 2-mercaptoethanesulfonate) and incubated for 8 hours at 37 °C followed by 8 hours at 4 °C to

cleave the intein and release the $Im7_{M1\text{-}S58}$ peptide thioester. To remove the CBD-intein (chitin binding domain), the mixture was then applied to a column of chitin resin of 0.5% the original culture volume. The flow through containing the desired $Im7_{M1\text{-}S58}$ peptide thioester was collected and purified by preparative reverse phase C_{18} HPLC, confirmed by ESI-MS to be within 1 Da of expected mass, checked by Coomassie staining of SDS-PAGE to be ≥95%, and lyophilized for storage. A 1 liter expression typically yields approximately 10 mg of desired product.

Native chemical ligation.

The appropriate lyophilized peptide pairs were dissolved in buffer P [100 mM sodium phosphate, pH 7] and quantified by UV (λ = 280 nm, ε = 9530 M^{-1} cm^{-1}). The synthetic glycopeptides was made to a concentration of 2 mM, while the expressed peptide was used in excess at 3 mM. To initiate the ligation reaction, an equal volume mixture of the appropriate peptides was mixed, followed by the addition of 2% benzyl metcaptan and 2% benzenelthiol. After 16 hours of shaking at room temperature, the precipitated thiols were removed by centrifugation (40,000 x *g*) and the full length Im7 variant was applied to a preparative reverse phase C_{18} HPLC, confirmed by ESI-MS to be within 1 Da of expected mass, checked by Coomassie staining of SDS-PAGE to be ≥95% pure and lyophilized for storage. By HPLC peak area, the typical yield for the ligation reaction was 80%.

Circular dichroism evaluation of Im7 variants.

Far-UV circular dichroism (CD) spectra were acquired on an Aviv Model 202 spectropolarimeter (Aviv Biomedical) using a 1 mm path length cell and a protein concentration of 250 μg/mL in buffer U [50 mM sodium phosphate, 400 mM sodium sulfate, 1 mM EDTA (ethylenediaminetetraacetic acid), pH 7] at 25°C.

Fluorescence evaluation of Im7 variants.

Fluorescence emission spectra of each Im7 variants were measured using a Photon Technology International Fluorimeter (Ford, West Sussex, UK). For spectra of native and denatured Im7 variants each protein was dissolved in native buffer U or denatured buffer U with 8 M urea to a protein concentration of approximately 50 μM. Excitation slit widths were set to 2 nm, emission slit widths were adjusted for protein concentration. Each spectrum was recorded from 270 nm to 450 nm in 1 nm increments, using an excitation wavelength of 280 nm. Spectra of all denatured states were assumed to have the same maximum intensity at 350 nm. The spectra of each native protein were normalized to the intensity of the respective denatured state, allowing direct comparison of the fluorescence intensity between variants.

References

1. Weerapana E, Imperiali B (2006) Asparagine-linked protein glycosylation: From eukaryotic to prokaryotic systems. *Glycobiology* **16**: 91-101.

2. Apweiler R, Hermjakob H, Sharon N (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta* **1473**: 4-8.

3. Meyer B, Moller H (2007) Conformation of glycopeptides and glycoproteins. *Top. Curr. Chem.* **267**: 187-251.

4. Burda P, Aebi M (1999) The dolichol pathway of N-linked glycosylation. *Biochim. Biophys. Acta.* **1426**: 239-57.

5. Betenbaugh MJ, Tomiya N, Narang S, Hsu JT, Lee YC (2004) Biosynthesis of human-type N-glycans in heterologous systems. *Curr Opin Struct Biol* **14**: 601-6.

6. Hamilton SR, Gerngross TU (2007) Glycosylation engineering in yeast: the advent of fully humanized yeast. *Curr Opin Biotechnol* **18**: 387-92.

7. Grogan MJ, Pratt MR, Marcaurelle LA, Bertozzi CR (2002) Homogeneous glycopeptides and glycoproteins for biological investigation. *Annu. Rev. Biochem.* **71**: 593-634.

8. Haase C, Seitz O (2007) Chemical synthesis of glycopeptides. *Top. Curr. Chem.* **267**: 1-36.

9. Liu L, Bennett CS, Wong CH (2006) Advances in glycoprotein synthesis. *Chem. Commun. (Camb.)* 21-33.

10. Martin R, Witte KL, Wong CH (1998) The synthesis and enzymatic incorporation of sialic acid derivatives for use as tools to study the structure, activity, and inhibition of glycoproteins and other glycoconjugates. *Bioorg. Med. Chem.* **6**: 1283-92.

11. Witte KL, Sears P, Martin R, Wong CH (1997) Enzymatic glycoprotein synthesis: preparation of ribonuclease glycoforms via enzymatic glycoprotein condensation and glycosylation. *J Am Chem Soc* **119**: 2114-2118.

12. Wang L, Brock A, Herberich B, Schultz PG (2001) Expanding the genetic code of Escherichia coli. *Science* **292**: 498-500.

13. Ryu Y, Schultz PG (2006) Efficient incorporation of unnatural amino acids into proteins in Escherichia coli. *Nat Methods* **3**: 263-5.

14. Liu H, Wang L, Brock A, Wong CH, Schultz PG (2003) A method for the generation of glycoprotein mimetics. *J. Am. Chem. Soc.* **125**: 1702-3.

15. Zhang Z, Gildersleeve J, Yang YY, Xu R, Loo JA, Uryu S, Wong CH, Schultz PG (2004) A new strategy for the synthesis of glycoproteins. *Science* **303**: 371-3.

16. Xu R, Hanson SR, Zhang Z, Yang YY, Schultz PG, Wong CH (2004) Site-specific incorporation of the mucin-type N-acetylgalactosamine-alpha-O-threonine into protein in Escherichia coli. *J. Am. Chem. Soc.* **126**: 15654-5.

17. Dawson PE, Muir TW, Clark-Lewis I, Kent SB (1994) Synthesis of proteins by native chemical ligation. *Science* **266**: 776-9.

18. Muir TW (2003) Semisynthesis of proteins by expressed protein ligation. *Annu. Rev. Biochem.* **72**: 249-89.

19. Perlstein MT, Atassi MZ, Cheng SH (1971) Desulfurization of sulfur amino acids and proteins with Raney nickel. *Biochim Biophys Acta* **236**: 174-82.

20. Pentelute BL, Kent SB (2007) Selective desulfurization of cysteine in the presence of Cys(Acm) in polypeptides obtained by native chemical ligation. *Org Lett* **9**: 687-90.

21. Wallis R, Leung KY, Pommer AJ, Videler H, Moore GR, James R, Kleanthous C (1995) Protein-protein interactions in colicin E9 DNase-immunity protein complexes. 2. Cognate and noncognate interactions that span the millimolar to femtomolar affinity range. *Biochemistry* **34**: 13751-9.

22. Dennis CA, Videler H, Pauptit RA, Wallis R, James R, Moore GR, Kleanthous C (1998) A structural comparison of the colicin immunity proteins Im7 and Im9 gives new insights into the molecular determinants of immunity-protein specificity. *Biochem J* **333 (Pt 1)**: 183-91.

23. Ferguson N, Capaldi AP, James R, Kleanthous C, Radford SE (1999) Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. *J Mol Biol* **286**: 1597-608.

24. Imperiali B, O'Connor SE (1999) Effect of N-linked glycosylation on glycopeptide and glycoprotein structure. *Curr. Opin. Chem. Biol.* **3**: 643-9.

25. O'Connor SE, Imperiali B (1996) Modulation of protein structure and function by asparagine-linked glycosylation. *Chem. Biol.* **3**: 803-12.

26. O'Connor SE, Imperiali B (1998) A molecular basis for glycosylation induced conformational switching. *Chem. Biol.* **5**: 427-437.

27. O'Connor SE, Pohlmann J, Imperiali B, Saskiawan I, Yamamoto K (2001) Probing the effect of the outer saccharide residues of N-linked glycans on peptide conformation. *J. Am. Chem. Soc.* **123**: 6187-8.

28. Bosques CJ, Tschampel SM, Woods RJ, Imperiali B (2004) Effects of glycosylation on peptide conformation: a synergistic experimental and computational study. *J. Am. Chem. Soc.* **126**: 8421-5.

29. Perler FB (2006) Protein splicing mechanisms and applications. *IUBMB Life* **58**: 63.

30. Hackenberger CP, O'Reilly MK, Imperiali B (2005) Improving glycopeptide synthesis: a convenient protocol for the preparation of beta-glycosylamines and the synthesis of glycopeptides. *J. Org. Chem.* **70**: 3574-8.

31. Likhosherstov L, Novikova O, Shibaev V (2002) New efficient synthesis of β-glucosylamines of mono and disaccharides with the use of ammonium carbamate. *Doklady Chem* **383**: 89-92.

32. Hackenberger CP, Friel CT, Radford SE, Imperiali B (2005) Semisynthesis of a glycosylated Im7 analogue for protein folding studies. *J. Am. Chem. Soc.* **127**: 12882-9.

33. Nagalingam A, Radford SE, Warriner S (2007) Avoidance of epimerization in the synthesis of peptide thioesters using Fmoc protection. *Synlett* **16**: 2517-2520.

34. Mathys S, Evans TC, Chute IC, Wu H, Chong S, Benner J, Liu XQ, Xu MQ (1999) Characterization of a self-splicing mini-intein and its conversion into autocatalytic N- and C-terminal cleavage elements: facile production of protein building blocks for protein ligation. *Gene* **231**: 1-13.

35. New England Biolabs Inc. (2004) Impact-TWIN: purification, ligation and cyclization of recombinant proteins using self-cleavable affinity tags. *Instruction Manual*

36. Hackenberger CP, Chen MM, Imperiali B (2006) Expression of N-terminal Cys-protein fragments using an intein refolding strategy. *Bioorg. Med. Chem.*

37. Spence GR, Capaldi AP, Radford SE (2004) Trapping the on-pathway folding intermediate of Im7 at equilibrium. *J Mol Biol* **341**: 215-26.

38. Capaldi AP, Kleanthous C, Radford SE (2002) Im7 folding mechanism: misfolding on a path to the native state. *Nat. Struct. Biol.* **9**: 209-16.

CHAPTER 3

EFFECT OF GLYCOSYLATION ON IM7 FOLDING KINETICS

Acknowledgements

The author is grateful to Alice Bartlett and Dr. Gareth Morgan from the University of Leeds for their instructions on the use of the stopped-flow fluorometer and data analysis.

In 1961, Anfinsen and co-workers demonstrated that denatured ribonuclease A (RNase A) spontaneously refolds to its native state without the requirement for other biological machinery (1), leading to the understanding that the precise three-dimensional structure of a protein is encoded within its amino acid sequence (2). Since this discovery, much effort has been devoted to understanding the mechanism of protein folding, including the major thermodynamic and kinetic forces that drive this process and the possibility of predicting structure from sequence (3,4). In addition to these important questions, the study of protein folding has also been motivated by the realization that the misfolding and aggregation of particular proteins is associated with a wide range of debilitating human diseases (5,6), such as cystic fibrosis (7), cataracts (8), bovine spongiform encephalopathy (mad-cow) (9), type II diabetes (10), and Alzheimers and Parkinsons diseases (11).

Unlike most types of protein modification, N-linked glycosylation in eukaryotics cells is a co-translational process that occurs as a nascent protein is translocated into the lumen of the endoplasmic reticulum (ER) before the entire protein has been biosynthesized by the ribosome (12). Since the glycan is installed on the acceptor protein before the protein reaches its native structure, N-linked glycosylation has long been implicated in the protein folding process. Indeed, there have been numerous examples in which glycoproteins are unable to attain their native structures without the appropriate N-linked glycan, a situation often found when mammalian glycoproteins are expressed in *Escherichia coli* (13,14). Since approximately 50% of all eukaryotic proteins carry N-linked glycans (15), it is important to understand the specific effects of the N-linked glycan on the protein folding process. Our current understanding of N-linked glycosylation reveals two different general mechanisms through which the glycans can

influence protein folding: through interactions with chaperone proteins and through direct structural effects on the local polypeptide backbone.

In vivo studies have found numerous quality controls and folding efficiency enhancement mechanisms within the ER that are N-linked glycan dependent. Many chaperone proteins within the ER are lectins that recognize and bind to specific features of the N-glycan on newly synthesized proteins (16). One well studied example of N-glycan mediated quality control involves the calnexin / calreticulin (CNX / CRT) pair of ER resident lectins, which bind to and facilitate the folding of proteins with the monoglucosylated glycan $Glc_1Man_9GlcNAc_2$ (17). The progress of the folding is checked by a glycosidase and glycosyltransferase pair, which removes the single remaining glucose if the protein is folded, or retains the glucose if the protein is still unfolded and requires further interaction with CNX and CRT. Glycoproteins undergoing futile folding attempts are trimmed of their terminal mannose residues to $Man_8GlcNAc_2$ or $Man_5GlcNAc_2$ by α-1,2-mannosidases within the ER, which target these misfolded proteins for degradation. The specific structure of the N-glycan provides a handle for the binding of chaperones and allows them to sense whether the glycoprotein is properly folded (17).

Both in vitro and in silico studies also point toward a more direct mechanism through which N-linked glycosylation can influence protein folding. The tetradecasaccharide glycan initially transferred to the acceptor protein is a large and hydrophilic moiety comparable in volume and surface area to the protein itself, especially since many proteins carry multiple glycans. Studies on glycopeptides reveal a dramatic conformational change from an extended Asx-turn to a more compactly folded type 1 β-turn upon glycosylation (**Figure 3-1**) (18). This structural change

does not appear to arise from hydrogen bond interactions between the glycan and the peptide, but rather through steric crowding or solvent-mediated effects (19). Furthermore, the effects of the glycan are heavily dependent on its specific structure. Substitution of the inner two N-acetyl glucosamines with glucose (20), as well as changing the native β-asparagine linkage to an α-linkage (21) both fail to induce β-turn turn formation. It has been proposed that the presence of an N-linked glycan in critical locations within the protein may promote folding by restricting the rotational backbone dihedral angles and overall conformational space available to the local structure (22,23).

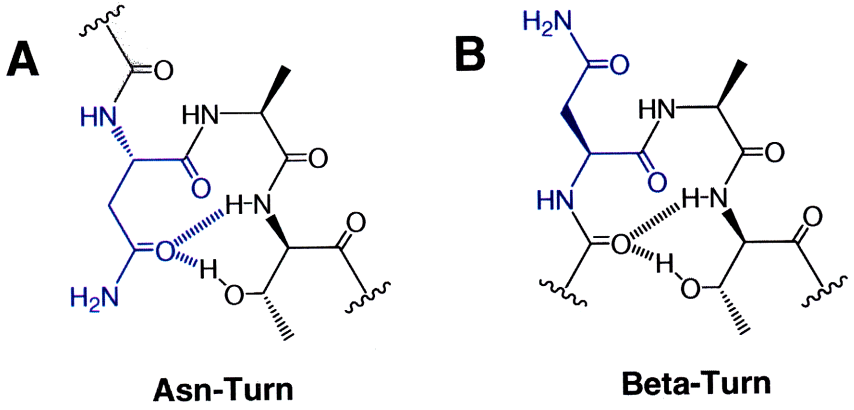


Figure 3-1. Example of **A**) an Asx-turn and **B**) a β-turn. The curvy lines represent the extension of the polypeptide.

Although the structural and thermodynamic effects of N-linked glycans on protein structure have been studied in many systems, there is comparatively little information on how N-linked glycans affect the kinetics of protein folding. This is primarily because of the difficulties in obtaining homogeneously glycosylated protein samples (24). Naturally glycosylated proteins have been isolated and used to demonstrate that refolding is much more efficient compared to the apo form. For example when the N-linked glycoprotein ECorL from *Erythrina corallodendron* was

refolded from 6 M urea the recovery was 80-90% while the apo form yielded only 10-20% from the same refolding protocol (25,26). However, the glycans on these glycoproteins are too heterogeneous for a quantitative kinetic analysis, and furthermore, it is not possible to modulate the glycan structure or location. Alternatively, chemically glycosylated proteins have also been used for these studies because of the ability to control the size and degree of glycosylation (27). However, the glycans used in these studies are not biologically relevant and the glycoproteins produced are still heterogeneous at a molecular level.

In this study we investigated the effects of N-linked glycans on the rates of protein folding, using protein samples that are homogeneously glycosylated with a biologically relevant chitobiose (GlcNAc-β-1,4-GlcNAc) glycan (28). This disaccharide is common to all eukaryotic N-glycans and has been previously found to have a significant effect on the conformation of polypeptides (20). Our model protein is a small 87 amino acid, single domain, 4-helix bacterial protein called immunity protein 7 (Im7), which is not natively glycosylated (29). Im7 is an ideal model protein for the study of protein folding because it has a well-characterized folding mechanism and lacks disulfide bonds, cis-proline amide bonds, and prosthetic groups that can complicate folding kinetics (30). It also has a fortuitous W75 residue which is fluorescent when solvent exposed in the unfolded state, and highly quenched in the native folded state when it is placed in close proximity to H47. This allows the folding of Im7 to be monitored by stopped flow fluorescence spectroscopy (30).

Using the semi-synthetic strategy as previously described in Chapter 2 (28), the asparagine-linked chitobiose glycan was site-specifically introduced into seven locations of different

secondary structural motifs within the Im7 sequence (**Figure 3-2**). Glycosylation sites 20 and 73 are within the middle of α-helices, sites 13 and 73 are close to the termini of these α-helices, sites 5 and 60 are within flexible loop regions, and glycosylation site 27 is located in a loop region following the end of helix I, a common motif found in N-linked glycoproteins. All the sites were chosen to be solvent-exposed in order to minimize any obvious steric clashes that would prevent the protein from folding to its native structure. This systematic placement of N-glycans at sites of interest will allow us to thoroughly probe the effects of the glycans and determine whether such effects are general or specific to the local secondary structure. Unlike studies involving native glycoproteins, our semi-synthetic strategy allows us to look beyond native glycosylation sites and focus on specific residues within locations of interest.

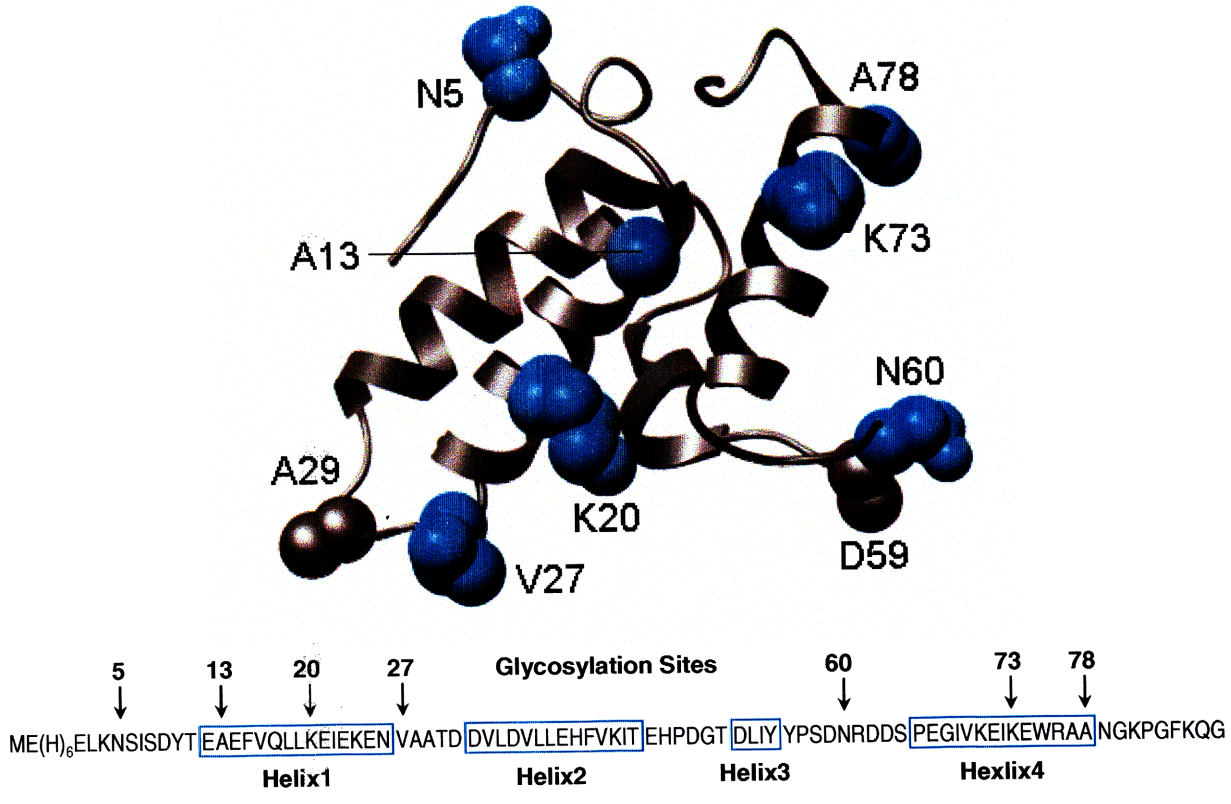


Figure 3-2. Ribbon diagram of Im7 (pdb code 1AYI) with ligation (dark) and glycosylation (blue) sites highlighted with spacefill side-chain representation. Amino acid sequence of Im7 is shown below with the α-helices marked by boxes.

3-1. Measurements of Im7 folding kinetics

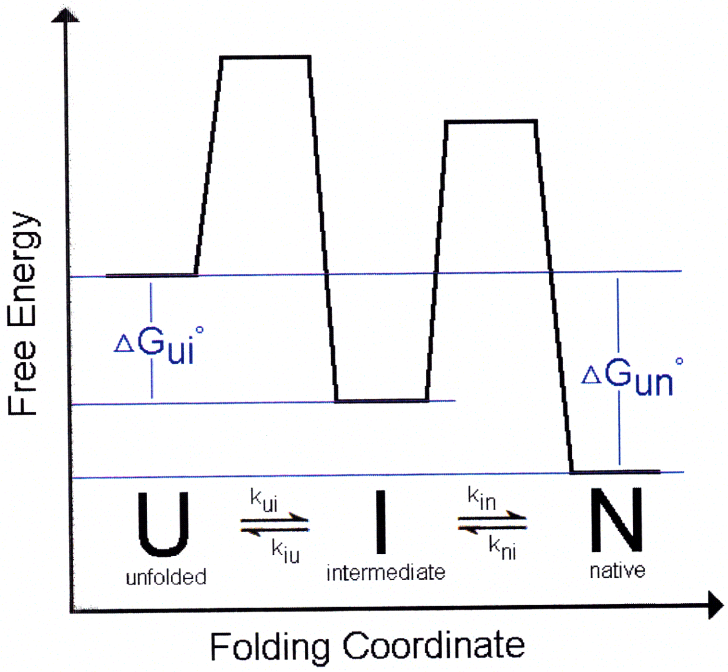


Figure 3-3. Free energy diagram of Im7 three-state folding pathway with arbitrary free energy scale. This linear progression from the unfolded, to the intermediate, and then the native state is referred to as an on-pathway three-state model. The rate constants are experimentally determined while the ΔG° values are calculated from the rate constants.

The folding of Im7 and its variants was studied using a stopped-flow fluorometer and procedures adapted from previous studies on Im7 (31). To measure the rate of refolding, each protein sample was dissolved in buffer at pH 7 containing 8 M urea denaturant and rapidly diluted with buffers of varying urea concentration from 0.75 M to 8 M. To measure the rate of unfolding, each sample was dissolved in buffer containing 0 M urea, and rapidly mixed with buffer containing 3 to 8 M urea. The fluorescence of W75 was continuously monitored during the mixing. Unlike many small single domain proteins, Im7 is known to fold by a three-state mechanism through a highly populated on-pathway intermediate (**Figure 3-3**) (32), consisting of three of the four native helices I, II, and IV and is stabilized by both native and non-native interactions (33). The intermediate is formed within the instrument deadtime (5 μs), making the

"U" to "I" transition too fast to be measured using our instrument. Therefore only the transition from the intermediate to the native folded state can be directly measured (31). This means the fluorescence signal can be fit to a single exponential function to determine the observed rate constant (k_{obs}) along with the initial and endpoint fluorescence.

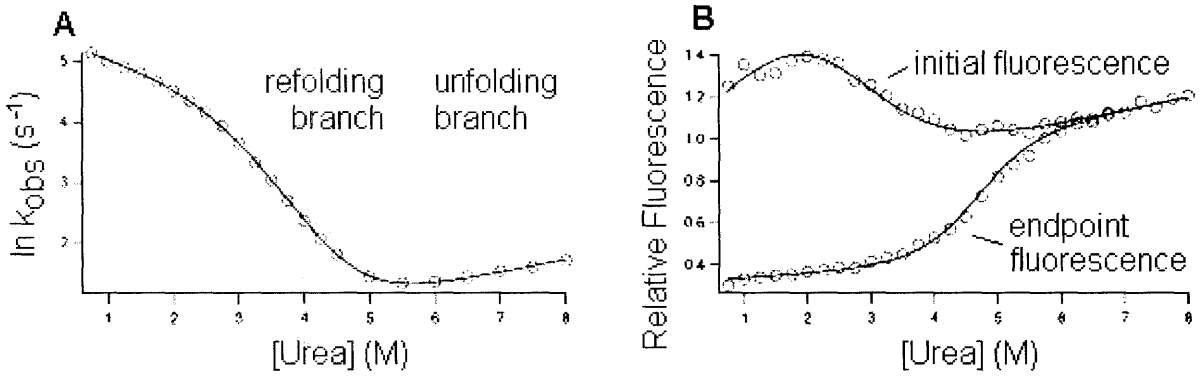


Figure 3-4. Wildtype Im7: **A**) Chevron plot and **B**) initial and endpoint fluorescence plot. The circles represent experimental data while the solid line represents the best fit to the on-pathway three-state model.

Plotting the natural log of the k_{obs} as a function of final urea concentration produces a Chevron plot (32,34), which consists of a refolding branch at low concentrations of urea, an unfolding branch at high concentrations of urea, and an inflection point in the middle corresponding to the urea concentration at which the unfolded and the native structures are equal in stability (**Figure 3-4**). The Chevron plot for each variant can be globally fitted along with the normalized initial and end point fluorescence data to an on-pathway three-state model (31). However, since the intermediate is formed in the deadtime of the instrument, we fixed the k_{ui} of each variant to be that of the wildtype Im7 previously measured using an ultra-fast mixing instrument. The other three rate constants were obtained directly from the Chevron plot global fitting. The M-values

represent the compactness of the overall protein, and were also obtained directly from the Chevron plot. From this mathematical analysis, we can obtain the kinetic and thermodynamic parameters for each Im7 variant, producing a complete folding kinetic characterization for each glycosylation site (**Table 3-1**).

3-2. Comparison of Im7 glyco-variants

Analysis of Im7 variants primarily involved comparing each glyco-variant to its corresponding non-glyco pseudo-wildtype control (**Figure 3-5**). This allows one to determine the precise effect of the chitobiose glycan on the kinetic and thermodynamic properties of Im7, as the two variants share the same peptide sequence and differ only by the presence of the glycan.

Table 3-1. The kinetic and thermodynamic parameters of each Im7 variant obtained from the global fitting of each Chevron plot and normalized initial and endpoint fluorescence data to an on-pathway three-state model (31). $K_{ui} = k_{ui}/k_{iu}$ where k_{ui} is estimated to be a constant 1573 s^{-1}. M_{un} is the global m-value, an indication that each Im7 variant has the same overall compactness. $\Delta G_{ui} = -RT\ln(K_{ui})$ and $\Delta G_{un} = -RT\ln(K_{ui}\, k_{in}/k_{ni})$.

	K_{ui}	k_{in} (s^{-1})	k_{ni} (s^{-1})	M_{un} (kJ mol^{-1} M^{-1})	ΔG_{ui} (kJ mol^{-1})	ΔG_{un} (kJ mol^{-1})
Wildtype	103 ± 10	271 ± 6	0.99 ± 0.05	5.1 ± 0.1	-10.9 ± 0.2	-24.1 ± 0.3
A29C N5	302 ± 58	151 ± 3	1.09 ± 0.11	4.7 ± 0.1	-13.4 ± 0.5	-25.0 ± 0.5
A29C N5-Glyco	352 ± 66	207 ± 4	1.07 ± 0.11	4.9 ± 0.1	-13.8 ± 0.4	-26.2 ± 0.5
A29C A13N	256 ± 68	198 ± 6	1.09 ± 0.10	5.3 ± 0.1	-13.1 ± 0.6	-25.3 ± 0.7
A29C A13N-Glyco	198 ± 115	179 ± 11	1.40 ± 0.22	5.3 ± 0.3	-12.7 ± 1.2	-24.1 ± 1.3
A29C K20N	68 ± 13	191 ± 7	1.43 ± 0.08	5.3 ± 0.1	-9.9 ± 0.4	-21.4 ± 0.5
A29C K20N-Glyco	15 ± 3	170 ± 12	2.58 ± 0.09	5.4 ± 0.4	-6.3 ± 0.4	-16.2 ± 0.5
A29C V27N	68 ± 8	166 ± 4	1.33 ± 0.05	5.2 ± 0.1	-9.9 ± 0.3	-21.3 ± 0.3
A29C V27N-Glyco	432 ± 93	140 ± 4	1.43 ± 0.13	5.3 ± 0.1	-14.3 ± 0.5	-25.1 ± 0.6
D59C N60	267 ± 35	159 ± 3	1.55 ± 0.09	5.1 ± 0.1	-13.2 ± 0.3	-24.1 ± 0.3
D59C N60-Glyco	296 ± 51	199 ± 5	1.25 ± 0.09	5.4 ± 0.1	-13.4 ± 0.4	-25.3 ± 0.4
D59C K73N	112 ± 16	134 ± 2	2.50 ± 0.10	4.9 ± 0.1	-11.1 ± 0.3	-20.5 ± 0.4
D59C K73N-Glyco	16 ± 5	132 ± 8	2.61 ± 0.10	5.1 ± 0.4	-6.5 ± 0.7	-15.7 ± 0.8
D54C A78N	92 ± 9	170 ± 3	1.62 ± 0.05	5.1 ± 0.1	-10.6 ± 0.2	-21.6 ± 0.2
D54C A78N-Glyco	64 ± 9	133 ± 4	1.28 ± 0.07	5.0 ± 0.1	-9.8 ± 0.3	-20.7 ± 0.4

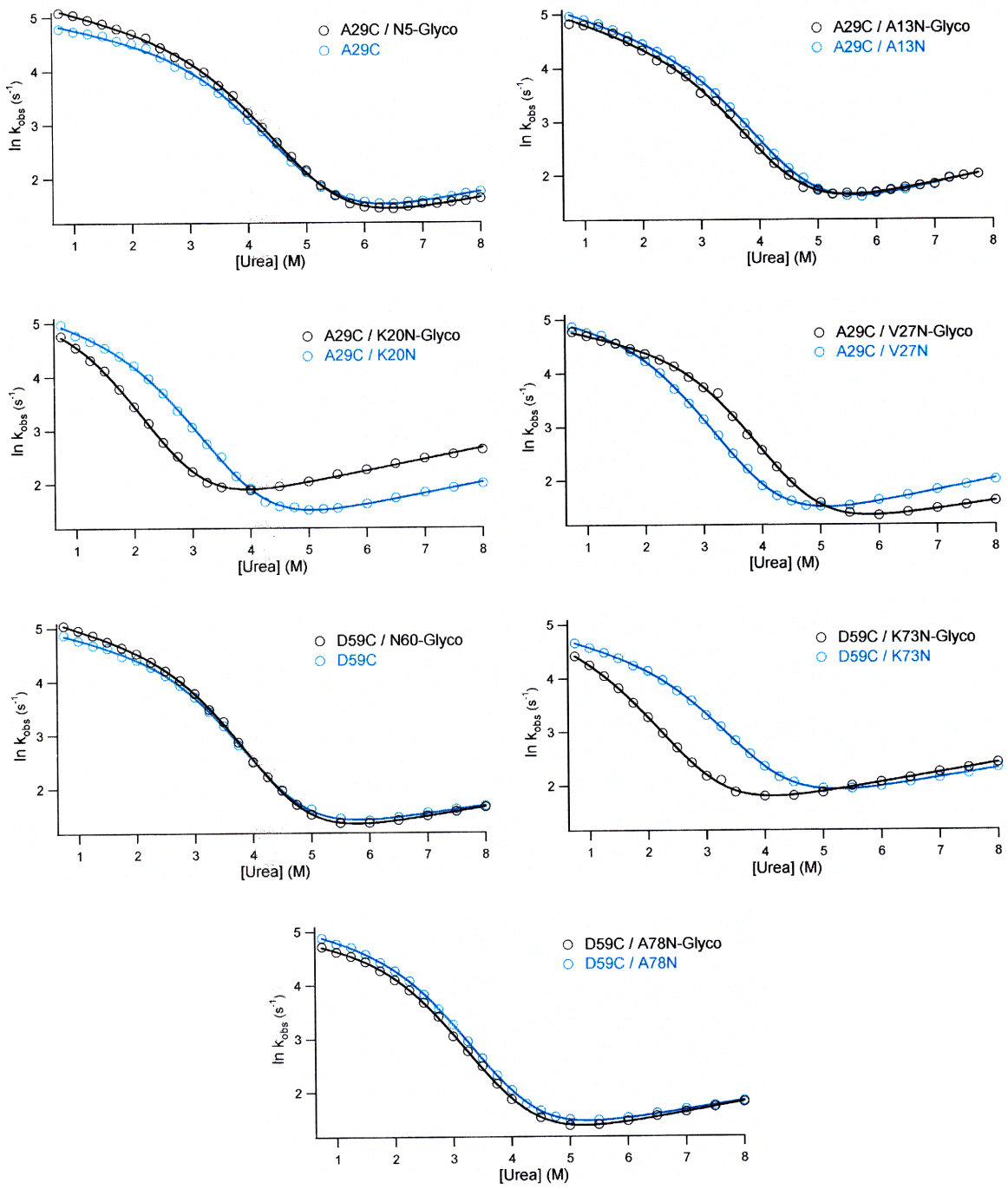


Figure 3-5. Chevron plots of each Im7 glyco-variant (black) compared with the corresponding non-glyco pseudo-wildtype control (blue). The circles represent experimental data while the solid line represents the best fit to the on-pathway three-state model.

Interestingly, the presence of the glycan can significantly affect both the rate of folding as well as the rate of unfolding, and this effect is highly dependent on the local structure of the glycosylation site (**Figure 3-2**). When the glycan was placed within a flexible loop region, such as in variants N5 and N60, the presence of the glycan did not appear to affect the folding of Im7 (**Figure 3-6**). Similarly, when the glycan was placed near the N- or C-terminus of an α-helix, such as in variants N13 and N78 respectively, the folding of Im7 was not perturbed. When the glycan was placed in the middle of a helix, such as in variants N20 and N73, folding is adversely affected signifying a less stable protein. However, the effect was not entirely the same for both variants, as both the rate of folding and unfolding is changed in the N20 variant, but only the rate of folding is affected in N73. In contrast, when the glycan was placed at a junction following the end of helix I at position N27, the folding rates were increased while the unfolding rates decreased, signifying an overall more stable protein.

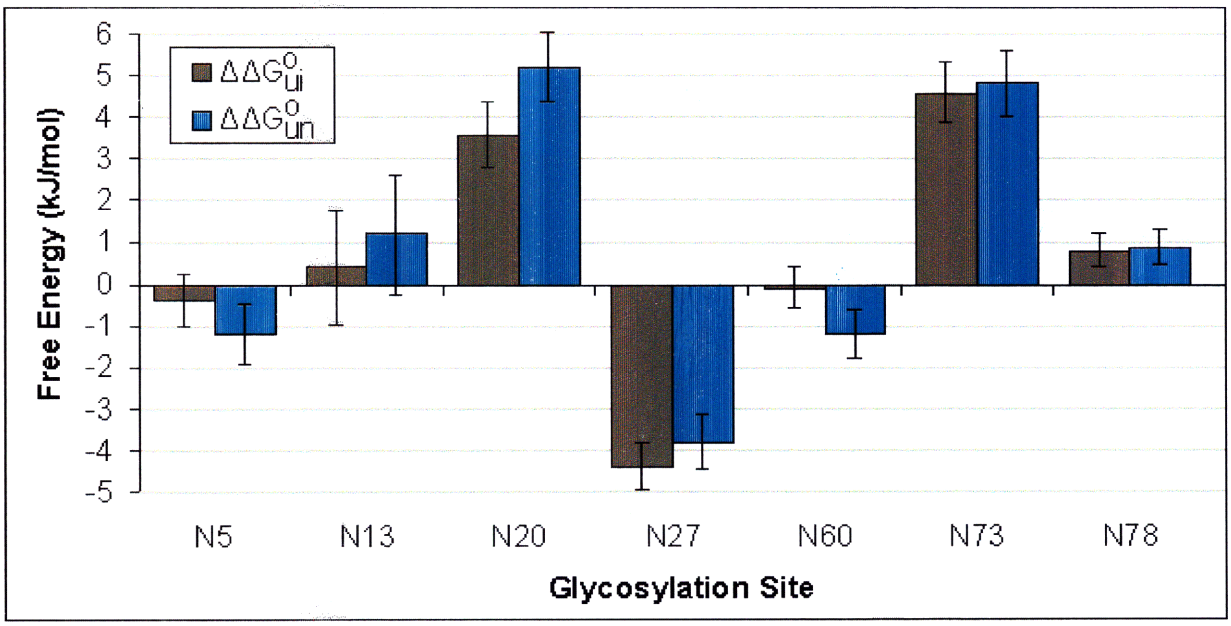


Figure 3-6. Comparison of the differences in the changes in free energy between the glyco-variants and their corresponding non-glyco pseudo-wildtypes. Actual values come from **Table 6-1**. $\Delta\Delta G^{o}_{ui} = \Delta G^{o}_{ui}(\text{glyco}) - \Delta G^{o}_{ui}(\text{non-glyco})$; $\Delta\Delta G^{o}_{un} = \Delta G^{o}_{un}(\text{glyco}) - \Delta G^{o}_{un}(\text{non-glyco})$.

Although the absolute differences between the glycol- and non-glyco-variants are small, they are significant relative to the overall stability of the protein. The total difference in free energy between native and unfolded Im7 is only 24 kJ/mol. A 5 kJ/mol change in free energy as a result of glycosylation represents more than 20% of the overall protein stability.

Despite the changes in folding kinetics observed, each glyco-variant of Im7 was still believed to fold to the same native structure for the following reasons: Each glycosylation site was chosen to be solvent exposed and contain no obvious interactions with its side chain to other residues. Each Im7 variant was characterized by circular dichorism and shown to exhibit the same degree of α-helical secondary structure as the wildtype (Chapter 2). Like wildtype Im7, each variant was observed to fold through a hyper-fluorescent intermediate within the first 5 μs of refolding, and its Chevron plots fit well to an on-pathway three-state model. Most importantly, fluorescence of W75 was highly quenched in the folded state of each variant (Chapter 2), which requires precise relative positioning of W75 with H47, consistent with a native fold. Therefore, when examining ΔG°_{un}, we make the assumption that the absolute free energy of the native state remained the same among each variant, and any changes in free energy occurred as a result of changes in the stability of the unfolded state.

3-3. Im7 N5 and N60 variants

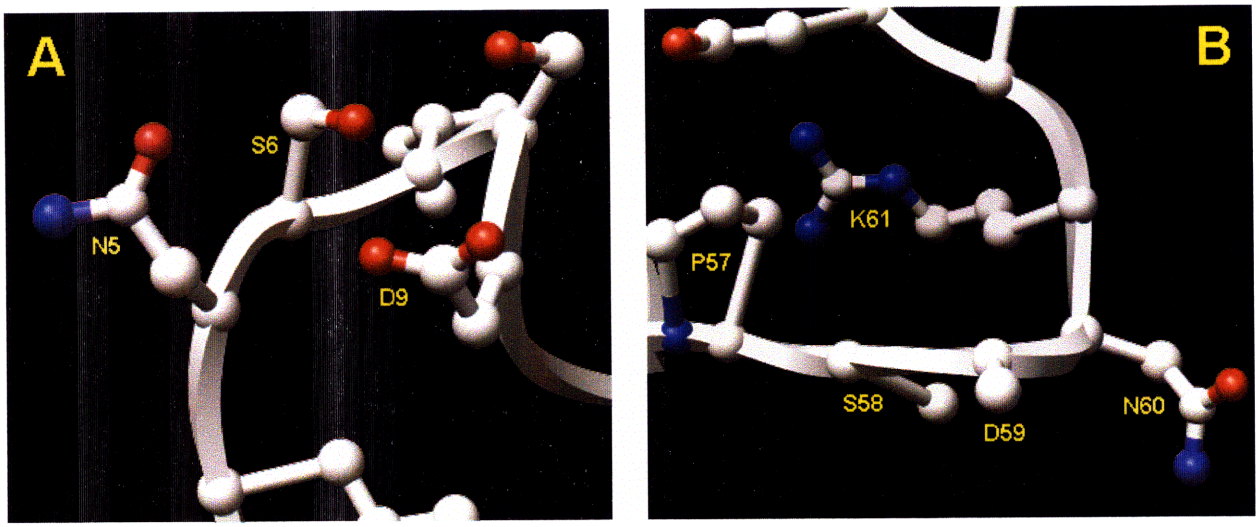


Figure 3-7. Crystal structure highlighting glycosylation sites N5 and N60 (pdb 1AYI).

The glycosylation sites at residues 5 and 60 (as well as 27) were chosen because they are located within loop regions of the protein (**Figure 3-7**). In a survey of 386 non-redundant glycosylation sites from protein structures found in the RCSB Protein Data Bank, approximately 40% of N-glycosylation occurred within loops, even though the overall incidence of loops within this data set was only 25% (22). It is important to note that neither site represents the dihedral angles of a β-turn, which has also been a commonly noted feature of eukaryotic N-link glycosylation. In any case, the frequency of N-linked glycosylation within loop regions prompted a closer look at the effect of placing an N-linked glycan within these structures.

Surprisingly, in both examples, the glycan exhibited little effect on the folding or unfolding properties of the protein. This may be due to the fact that both regions of the protein are flexible and solvent exposed, with no obvious nearby residues to interact with the glycan. Therefore, while the protein collapses to its intermediate state and rearranges to the native fold, the presence of the glycan can be accommodated without any noticeable changes in the folding process.

3-4. Im7 N20 and N73 variants

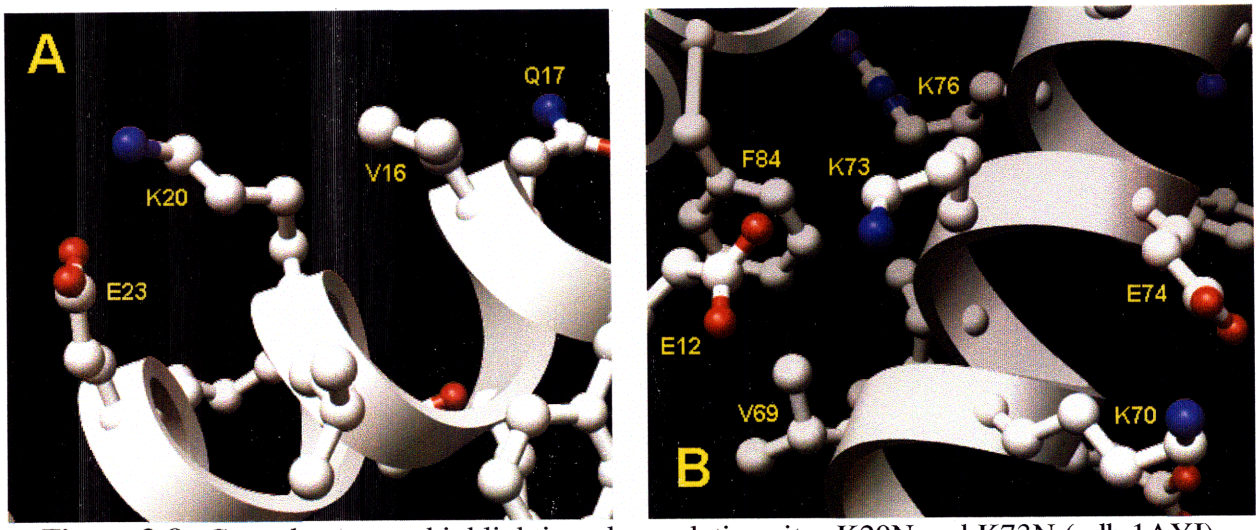


Figure 3-8. Crystal structure highlighting glycosylation sites K20N and K73N (pdb 1AYI).

The glycosylation sites at residues 20 and 73 are both located within the middle of α-helices (I and IV respectively) (**Figure 3-8**). Importantly, N-linked glycosylation is not frequently observed within α-helices. In the previous survey of known glycosylation sites, only 11% of N-linked glycans were found within helices, compared with an overall incidence of helical structure of 23% (22). In fact, α-helices represented the lowest incidence of N-linked glycosylation sites amongst all types of secondary structures. It is therefore interesting to examine the effect of glycosylation within these locations and understand the reason for this observation.

In both cases, the presence of the chitobiose disaccharide led to a destabilization of the overall protein by approximately 5 kJ/mol. Specifically, for both variants, the rate of refolding was significantly decreased relative to the corresponding pseudo-wildtypes. However, the rate of unfolding for the N20 variant was increased while unfolding for N73 remains roughly unchanged. The overall destabilization is likely to be caused by steric effects. Within the middle of an α-helix, each amino acid side chains is in close proximity to the side chains of residues

three and four amino acids away in both directions, which are located on the adjacent turns of the α-helix. The large size of the glycan attached to the asparagine side chain can cause it to clash with side chains of amino acids one turn above and below. This steric interaction may stabilize the unfolded state of Im7 relative to the native state since a random coil conformation is better able to accommodate the steric bulk of the glycan as compared to an α-helix.

It is unclear exactly why the changes in unfolding rates for the two variants differ. However, upon closer examination, it appeared that the two Chevron plots for the N20 and N73 glyco-variants were very similar, and rather it was the unfolding rate for the N73 pseudo-wildtype that was much higher than the other pseudo-wildtypes. This may be caused by additional unfavorable interactions caused by the K73N mutation, such as non-native bonding with nearby residues in the native or unfolded state, which is eliminated once N73 becomes glycosylated.

3-5. Im7 N13 and N78 variants

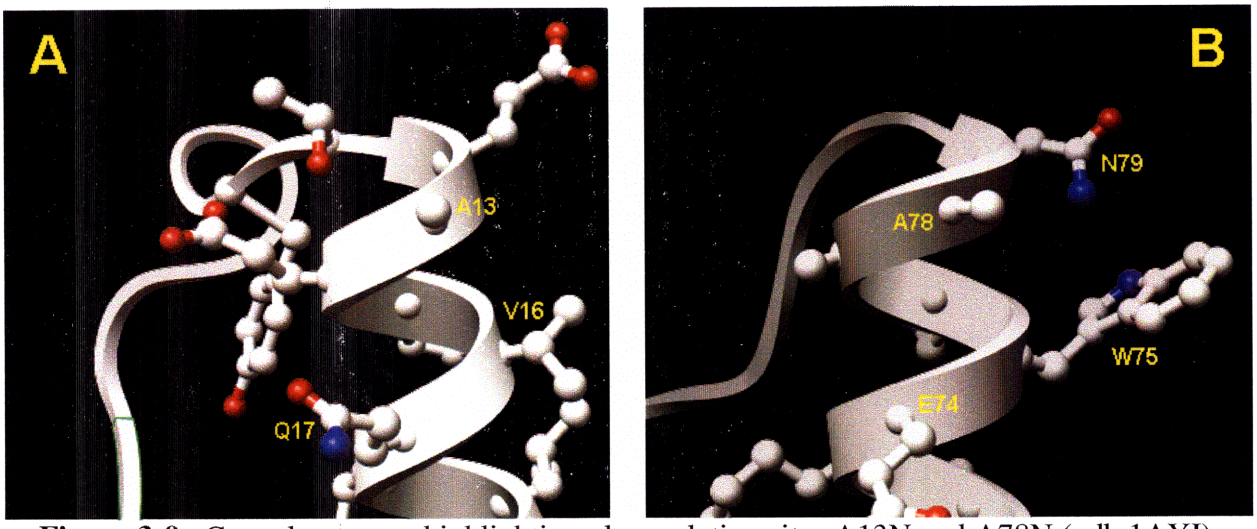


Figure 3-9. Crystal structure highlighting glycosylation sites A13N and A78N (pdb 1AYI).

Both residues N13 and N78 are also found within α-helices (I and IV respectively) (**Figure 3-9**). However, rather than at the center of a helix, N13 is located at the N-terminus and N78 is at the C-terminus of the respective helices. In this case, the asparagine chitobiose side chain, rather than being sandwiched between two amino acid side chains above and below, has significantly more space to occupy and is therefore better able to avoid the unfavorable steric clashes experienced by the N20 and N73 glyco-variants in the middle of the same helices. As a result, glycosylation at both the N13 and N78 sites did not exhibit significant changes in their folding kinetics.

3-6. Im7 N27 variant

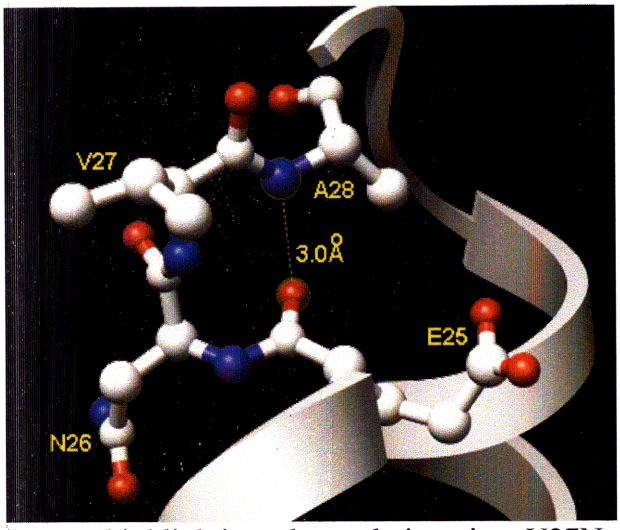


Figure 3-10. Crystal structure highlighting glycosylation sites V27N and the type-1 β-burn at that position (pdb 1AYI)

The N27 glyco-variant is unique within our set of Im7 variants because it was the only variant that appears to significantly promote folding. The placement of the chitobiose glyan at N27 increased the rate of folding while decreasing the rate of unfolding, indicating an overall more stable protein by approximately 4.5 kJ/mol in comparison to its relevant pseudo-wildtype control (**Figure 3-6**). Residue 27 is located within the loop region between helices I and II, and is the

first residue within this loop following the end of helix I (**Figure 3-10**). Interestingly, it has been observed in statistical analysis of N-linked glycoprotein that there is a highly increased probability of glycosylation site occurring at or just after points in the peptide chain where there is a change in secondary structure (22).

There are a number of effects that the glycan can impart at this critical position that would promote folding. Firstly it may promote efficient folding by acting as a marker for the end of helix I. It has been previously demonstrated within glycopeptides that N-linked glycans often lead to a more compact β-turn structure through steric crowding (**Figure 3-1**) (20). In this case, residue 27 does correspond to position "i + 1" of a type-1 β-turn. Therefore, it is possible that placing the chitobiose glycan at position 27 simply help terminates helix I and promote the formation of a β-turn necessary within this loop region. Alternatively, it has also been shown that N-glycans can promote folding by reducing the conformational freedom of the local peptide backbone and thus reduce the loss of configurational entropy on folding (35). This is also a plausible scenario since N27 is within a flexible loop which pays an entropic penalty when it becomes structured in the native state. It is likely that both of these two factors promote the folding of this Im7 variant.

Molecular dynamics simulations of short peptides corresponding to each glycosylation site are currently being carried out. Because of their flexibility, the peptides should provide an appropriate model for the unfolded state. By comparing the conformations of the peptides with and without the glycan, we can more closely examine the effects of the glycan on the local backbone structure.

Conclusion

To study the effects of N-linked glycosylation on protein folding, we have developed, for the first time, a unique system involving a small, well characterized model protein, Im7, which is homogeneously glycosylated with a biologically relevant chitobiose disaccharide glycan at seven carefully selected sites of differing secondary structure. Like the wildtype Im7, each glycoprotein variant appeared to fold through a well-populated folding intermediate and to the same native folded structure as determined by fluorescence spectroscopy and circular dichroism.

Using stopped-flow fluorescence spectroscopy, the folding kinetics of each Im7 glyco-variant was measured and compared with the corresponding non-glyco pseudo-wildtype control, which contained the same peptide backbone sequence and differed only by the presence of the glycan. The presence of the glycan can significantly affect both the rates of folding and unfolding, and the overall effect appears to be highly specific to the local structure of the glycosylation site. In summary, when the glycan was placed at a residue within a flexible loop or a non-sterically hindered location of a helix, the glycan is accommodated with minimal effects on protein folding rates. When the glycan was placed within the middle of a helix, where there can be possible steric clashes with adjacent side chains on the same face of the helix, protein folding is impeded most likely because the native structure is less able to accommodate the glycan in comparison to the more open unfolded state. Interestingly, protein folding was promoted by the placement of a glyan at a junction between a helix and the following turn, a common structural feature where N-glycans are found. This effect may be caused by both the rigidifying of the peptide backbone and the introduction of the turn to mark the end of the helix.

Experimental

Stopped flow kinetics measurements of Im7 variants.

Im7 variants of ≥95% purity were prepared as described in Chapter 5. Folding experiments were adapted from previously published studies on Im7 (31). Fluorescence measurements were carried out on an Applied-Photophysics SX1.8MV stopped-flow fluorimeter with the temperature held constant at 10 °C. For unfolding experiments, Im7 variants were dissolved at a concentration of approximately 0.35 mg/mL in buffer U [50 mM sodium phosphate, 400 mM sodium sulfate, 1 mM EDTA (ethylenediaminetetraacetic acid), pH 7] and rapidly mixed by 1:10 dilution into solutions consisting of buffer U with urea concentrations of 3.0-8.0 M, at 0.25 M increments. Refolding experiments were carried out in the same manner, except with Im7 variants dissolved in buffer U with 8 M urea, and mixed into solutions containing urea ranging from 0.75-8.0 M.

At each urea concentration at least 7 kinetic traces were averaged and fit to a single exponential function using the manufacturer's software in order to obtain the observed rate constant (k_{obs}) as well as the initial and final fluorescence signals. After subtracting the buffer blanks, the initial and final fluorescence signals were normalized to the fluorescence signal of the 7.75 M urea measurement. The k_{obs} and normalized endpoint and initial fluorescence signals for each variant were fit to the function of an on-pathway three-state model using Igor Pro 6.0 (Wavemetrics) as previously described (31). The k_{ui} and m_{ui} were fixed to the values previously obtained for wildtype Im7 from ultra-fast mixing continuous-flow measurements (k_{ui} = 1573 s^{-1}, m_{ui} = 1.23 kJ mol^{-1} M^{-1}). Intermediate stability was determined by allowing k_{iu} to vary (31).

References

1. Anfinsen CB, Haber E, Sela M, White FH, Jr. (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci U S A* **47**: 1309-14.

2. Anfinsen CB (1973) Principles that govern the folding of protein chains. *Science* **181**: 223-30.

3. Englander SW, Mayne L, Krishna MM (2007) Protein folding and misfolding: mechanism and principles. *Q Rev Biophys* **40**: 287-326.

4. Fersht AR (2008) From the first protein structures to our current knowledge of protein folding: delights and scepticisms. *Nat Rev Mol Cell Biol* **9**: 650-4.

5. Murphy RM (2002) Peptide aggregation in neurodegenerative disease. *Annu Rev Biomed Eng* **4**: 155-74.

6. Lee C, Yu MH (2005) Protein folding and diseases. *J Biochem Mol Biol* **38**: 275-80.

7. Cheung JC, Deber CM (2008) Misfolding of the cystic fibrosis transmembrane conductance regulator and disease. *Biochemistry* **47**: 1465-73.

8. Pande A, Pande J, Asherie N, Lomakin A, Ogun O, King J, Benedek GB (2001) Crystal cataracts: human genetic cataract caused by protein crystallization. *Proc Natl Acad Sci U S A* **98**: 6116-20.

9. Dalsgaard NJ (2002) Prion diseases. An overview. *Apmis* **110**: 3-13.

10. Hayden MR, Tyagi SC, Kerklo MM, Nicolls MR (2005) Type 2 diabetes mellitus as a conformational disease. *Jop* **6**: 287-302.

11. Selkoe DJ (2004) Cell biology of protein misfolding: the examples of Alzheimer's and Parkinson's diseases. *Nat Cell Biol* **6**: 1054-61.

12. Weerapana E, Imperiali B (2006) Asparagine-linked protein glycosylation: From eukaryotic to prokaryotic systems. *Glycobiology* **16**: 91-101.

13. Kelley F, Winkler M (1990) Folding of eukaryotic proteins produced in Escherichia coli. *Genet Eng* **12**: 1-19.

14. Arango R, Adar R, Rozenblatt S, Sharon N (1992) Expression of Erythrina corallodendron lectin in Escherichia coli. *Eur J Biochem* **205**: 575-81.

15. Apweiler R, Hermjakob H, Sharon N (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta* **1473**: 4-8.

16. Helenius A, Aebi M (2004) Roles of N-linked glycans in the endoplasmic reticulum. *Annu. Rev. Biochem.* **73**: 1019-49.

17. Caramelo JJ, Parodi AJ (2008) Getting in and out from calnexin/calreticulin cycles. *J Biol Chem* **283**: 10221-5.

18. O'Connor SE, Imperiali B (1997) Conformational Switching by Asparagine-Linked Glycosylation. *J. Am. Chem. Soc.* **119**: 2295-2296.

19. Imperiali B, O'Connor SE (1999) Effect of N-linked glycosylation on glycopeptide and glycoprotein structure. *Curr. Opin. Chem. Biol.* **3**: 643-9.

20. O'Connor SE, Imperiali B (1998) A molecular basis for glycosylation induced conformational switching. *Chem. Biol.* **5**: 427-437.

21. Bosques CJ, Tschampel SM, Woods RJ, Imperiali B (2004) Effects of glycosylation on peptide conformation: a synergistic experimental and computational study. *J Am Chem Soc* **126**: 8421-5.

22. Petrescu AJ, Milac AL, Petrescu SM, Dwek RA, Wormald MR (2004) Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology* **14**: 103-14.

23. Meyer B, Moller H (2007) Conformation of glycopeptides and glycoproteins. *Top. Curr. Chem.* **267**: 187-251.

24. Liu L, Bennett CS, Wong CH (2006) Advances in glycoprotein synthesis. *Chem. Commun. (Camb.)* 21-33.

25. Mitra N, Sinha S, Ramya TN, Surolia A (2006) N-linked oligosaccharides as outfitters for glycoprotein folding, form and function. *Trends Biochem Sci* **31**: 156-63.

26. Mitra N, Sharon N, Surolia A (2003) Role of N-linked glycan in the unfolding pathway of Erythrina corallodendron lectin. *Biochemistry* **42**: 12208-16.

27. Sola RJ, Rodriguez-Martinez JA, Griebenow K (2007) Modulation of protein biophysical properties by chemical glycosylation: biochemical insights and biomedical implications. *Cell Mol Life Sci* **64**: 2133-52.

28. Hackenberger CP, Friel CT, Radford SE, Imperiali B (2005) Semisynthesis of a glycosylated Im7 analogue for protein folding studies. *J. Am. Chem. Soc.* **127**: 12882-9.

29. Dennis CA, Videler H, Pauptit RA, Wallis R, James R, Moore GR, Kleanthous C (1998) A structural comparison of the colicin immunity proteins Im7 and Im9 gives new insights into the molecular determinants of immunity-protein specificity. *Biochem J* **333 (Pt 1)**: 183-91.

30. Ferguson N, Capaldi AP, James R, Kleanthous C, Radford SE (1999) Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. *J Mol Biol* **286**: 1597-608.

31. Morton VL, Friel CT, Allen LR, Paci E, Radford SE (2007) The effect of increasing the stability of non-native interactions on the folding landscape of the bacterial immunity protein Im9. *J Mol Biol* **371**: 554-68.

32. Brockwell DJ, Radford SE (2007) Intermediates: ubiquitous species on folding energy landscapes? *Curr Opin Struct Biol* **17**: 30-7.

33. Spence GR, Capaldi AP, Radford SE (2004) Trapping the on-pathway folding intermediate of Im7 at equilibrium. *J Mol Biol* **341**: 215-26.

34. Zarrine-Afsar A, Davidson AR (2004) The analysis of protein folding kinetic data produced in protein engineering experiments. *Methods* **34**: 41-50.

35. Hoffmann D, Florke H (1998) A structural role for glycosylation: lessons from the hp model. *Fold Des* **3**: 337-43.

CHAPTER 4

CHEMO-ENZYMATIC SYNTHESIS OF UNDECAPRENOL-LINKED SUBSTRATES

Acknowledgements.

The original chemical synthesis of UDP-bacillosamine was developed by Dr. Eranthie Weerapana from our lab, as published in:

Weerapana E, Glover KJ, Chen MM, Imperiali B (2005) Investigating bacterial N-linked glycosylation: synthesis and glycosyl acceptor activity of the undecaprenyl pyrophosphate-linked bacillosamine. *J. Am. Chem. Soc.* **127**: 13766-7.

Glover KJ, Weerapana E, Chen MM, Imperiali B (2006) Direct biochemical evidence for the utilization of UDP-bacillosamine by PglC, an essential glycosyl-1-phosphate transferase in the *Campylobacter jejuni* N-linked glycosylation pathway. *Biochemistry* **45**: 5343-50.

The cloning and purification of the UDP-bacillosamine biosynthetic enzymes PglF, E, and D were carried out by Dr. Nelson Olivier. The optimization of the reactions conditions for PglF, E, and D was also carried out by Dr. Olivier, although the conditions for the one-pot 8-enzyme synthesis of undecaprenyl diphosphate-heptasaccharide were determined by the author, as published in:

Olivier NB, Chen MM, Behr JR, Imperiali B (2006) *In Vitro* Biosynthesis of UDP-N,N'-Diacetylbacillosamine by Enzymes of the *Campylobacter jejuni* General Protein Glycosylation System. *Biochemistry* **45**: 13659-69.

The lectin binding studies were published in:

Reid CW, Stupak J, Chen MM, Imperiali B, Li J, Szymanski CM (2008) Affinity-capture tandem mass spectrometric characterization of polyprenyl-linked oligosaccharides: tool to study protein N-glycosylation pathways. *Anal Chem* **80**: 5468-75.

Asparagine-linked glycosylation has long been considered a unique feature of the eukaryotic and archaea kingdom. Recently, a gram-negative bacterium called *Campylobacter jejuni* became the first bacterial organism discovered to also carry out this process (1-3). The enzymes involved in this protein glycosylation pathway (Pgl) have since been identified and characterized (3-5). An oligosaccharide is built up by a series of glycosyltransferases (PglC, A, J, H, I) on a polyisoprenyl-carrier anchored to the periplasmic membrane (**Figure 4-1**). A flippase (PglK) then flips the fully assembled glycan from the cytoplasmic to the periplasmic side of the membrane where it is transferred by an oligosaccharyltransferase (PglB) from the polyisoprenol-carrier to the asparagine side chain of the acceptor protein. Although the polyisoprenol-carrier and the oligosaccharide are different to those in the eukaryotic pathways (**Figure 4-2**), the process and machinery are highly analogous (6).

Table 4-1. Comparison of eukaryotic and prokaryotic N-linked glycosylation processes (6).

Characteristic	**Eukaryotic** *(Saccharomyces cerevisiae)*	**Bacterial** *(Campylobacter jejuni)*
Location	Lumen of the ER	Bacterial periplasm
Timing	Co-translational	Post-translational
Glycosyl donor	Dolichyl diphosphate-linked tetradecasaccharide	Undecaprenyl diphosphate-linked heptasaccharide
Glycosylation sequence	N-X-S/T where $X \neq P$	D/E-X_1-N-X_2-S/T where $X \neq P$
Oligosaccharyl transferase	8 subunit complex	Single protein PglB
Glycan trimming and elaboration	In both the ER and Golgi	No further processing

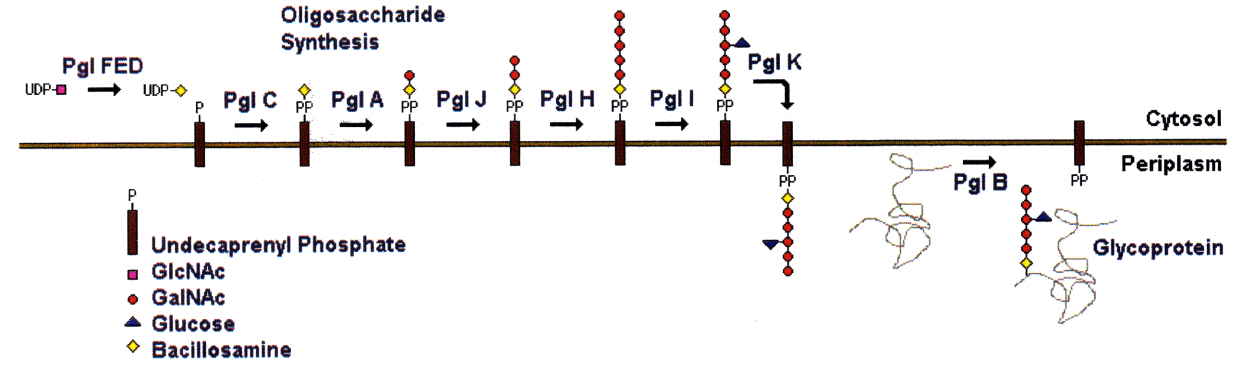


Figure 4-1. *C. jejuni* N-linked glycosylation pathway.

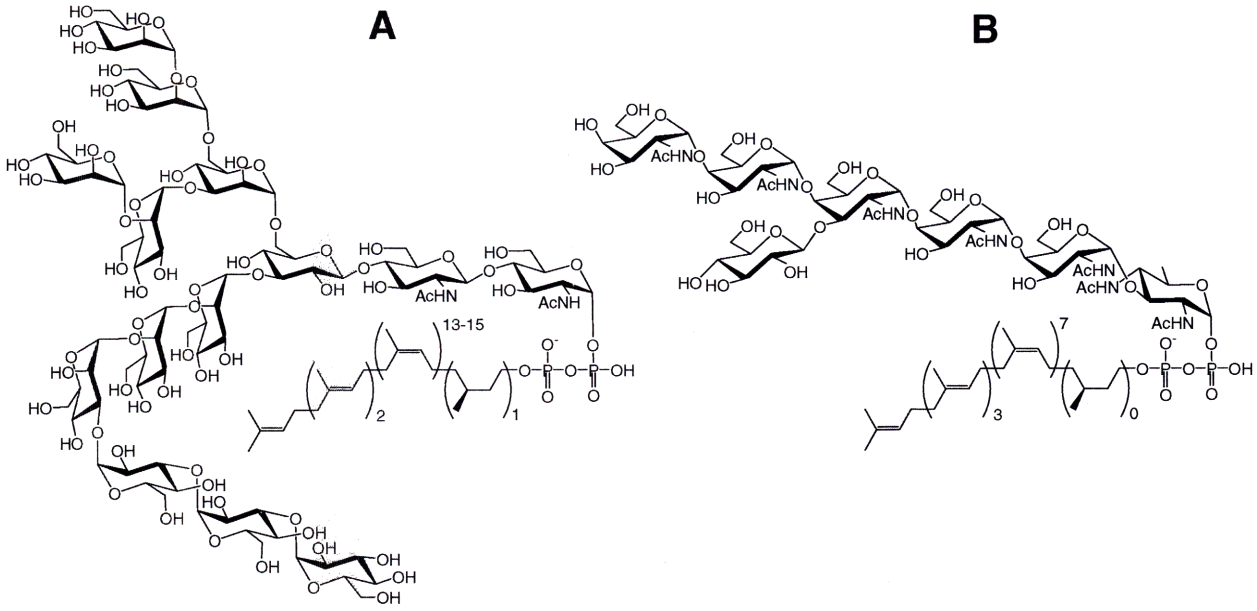


Figure 4-2. Structures of the *S. cerevisiae* and *C. jejuni* polyisoprenyl-linked glyan donors in comparison: **A**) dolichyl diphosphate-tetradecasaccharide. **B**) undecaprenyl diphosphate-heptasaccharide.

Although there has been much interest in the study of eukaryotic N-linked glycosylation, *C. jejuni* provides many unique opportunities and advantages as a model for understanding this complex process. Firstly, the *C. jejuni* pathway appears to be a more straightforward system with fewer enzymes and substrates. There are only five glycosyltransferases, which have all

been identified (3,4), whereas the eukaryotic system contains many more glycosyltransferases in order to assemble its larger glycan (14 sugars vs. 7). There are only three types of sugars involved and they are all activated as uridine diphosphate (UDP) derivatives, where as the eukaryotic system also employs membrane-bound dolichyl phosphate-linked sugar donors. Secondly, each enzyme within this pathway can be recombinantly overexpressed in *E. coli* (5), which is a more convenient and robust protein production system in comparison to yeast, and does not possess an interfering N-link glycosylation system of its own. Thirdly, the oligosaccharyltransferase (OT) in *C. jejuni* appears to be fully functional as a single subunit, which makes it a more promising target for mechanistic and crystallographic studies in comparison to its 8-subunit eukaryotic counterpart. Finally, *C. jejuni* is a food-borne pathogen, and it has been shown that the N-linked glycans play a critical role in host adherence, invasion, and colonization (7). Therefore, understanding the *C. jejuni* N-linked glycosylation pathway may facilitate the development of therapeutics towards the gastroenterological disorders caused by this pathogenic bacterium.

In order to study the individual enzymes of *C. jejuni* Pgl pathway in vitro, access to the substrates and substrate intermediates is required. Since *C. jejuni* is pathogenic and does not produce a large amount of glycan substrates, purification from native sources is both unsafe and impractical. To obtain large quantities of highly pure substrates, we employed a chemo-enzymatic synthesis, combining the versatility of organic synthesis with the specificity and selectivity of the enzymes from the *C. jejuni* Pgl pathway itself. This approach also offers the advantage of incorporating tritium [^{3}H]-radiolabels into the substrates, allowing for sensitive in vitro assays in the low micromolar concentration range and the synthesis of unnatural substrate

analogues. Access to these radiolabeled substrates, intermediates, and analogues is essential in allowing us to characterize and probe the enzymes of interest from this exciting new N-linked glycosylation pathway. For example, Chapters 5 and 6 will demonstrate the usefulness of the radiolabeled polyisoprenol-linked disaccharide produced in vitro for the evaluation of both the peptide acceptor and polyisoprenol-carrier specificity in the oligosaccharyltransferase PglB.

4-1. Chemical synthesis of UDP-bacillosamine

The first sugar to be transferred to the undecaprenyl-carrier is the highly modified 2,4-diacetamido-bacillosamine (2,4-diacetamido-2,4,6-trideoxy-β-D-glucopyranose), which will be referred to as "bacillosamine" in this dissertation. It is synthesized as the activated UDP sugar nucleotide donor, and then transferred by the glycosylphosphotransferase, PglC, to the undecapreyl phosphate-carrier (**Figure 4-3**) (8). This unusual sugar also appears to be an important feature of bacterial O-linked glycosylation, as its derivatives have been discovered in numerous other bacteria (9), including the O-linked pillin glycosylation system of *Neisseria gonorrhoeae* (10). In *C. jejuni*, bacillosamine is especially important since this sugar is present in all N-linked glycoproteins as well as in every undecaprenol-linked intermediate.

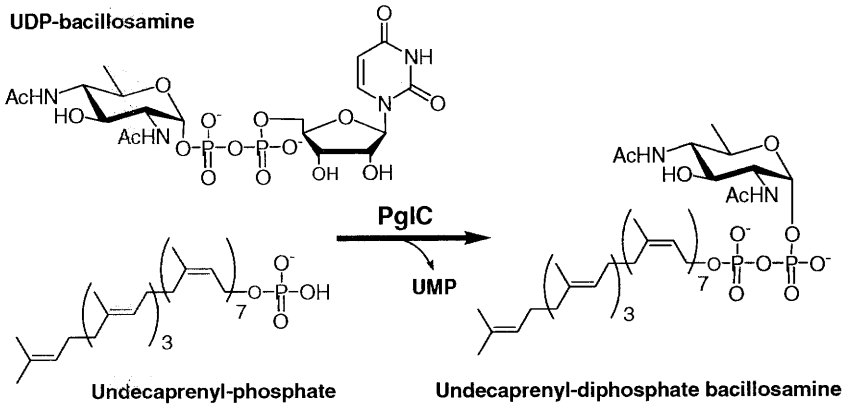
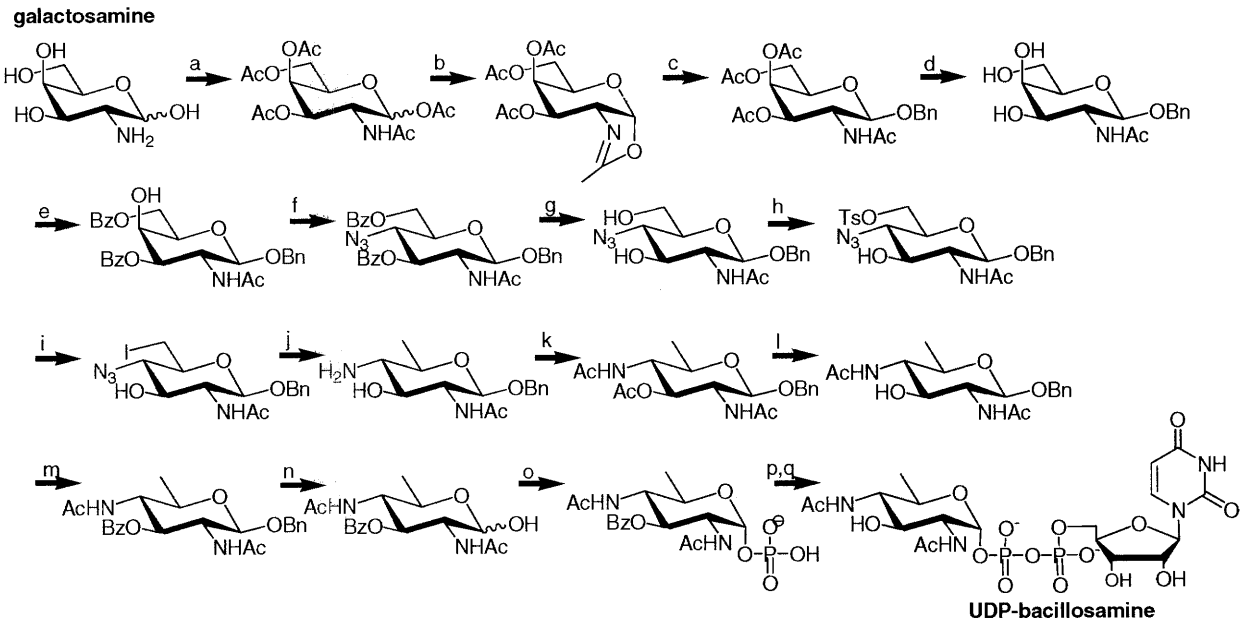


Figure 4-3. Biosynthesis of undecaprenyl diphosphate-bacillosamine using PglC.

Because UDP-bacillosamine is not commercially available, a chemical synthesis was designed and carried out by Dr. Eranthie Werapana in our laboratory. This multi-step synthesis allowed for the production of highly pure UDP-bacillosamine in milligram quantities starting from commercially available galactosamine (**Figure 4-4**) (8,11). However, the yield of this synthesis was modest, not only because the protocol was lengthy but also because the synthesis was entirely linear. From 10 g of galactosamine, approximately 15 mg of UDP-bacillosamine can be produced following the original protocol, corresponding to an overall yield of ~0.5%. Furthermore, chemical intermediates of sufficient purity for NMR characterization were difficult to obtain. To improve upon the yield and purification protocol of the original synthesis, a number of small adjustments were made to optimize each step, including two notable changes.



(a) Ac_2O, pyridine; (b) $FeCl_3$, CH_2Cl_2; (c) BnOH, Tos-OH, toluene, $MeNO_2$, 110°C; (d) NaOMe, MeOH; (e) BzCl, pyridine, -40 to 0°C; (f) i) Tf_2O, CH_2Cl_2/pyridine, 0°C to rt; ii) NaN_3, DMF; (g) NaOMe, MeOH; (h) TsCl, pyridine, 0°C to rt; (i) NaI, MeCN, 80°C; (j) H_2, $Pd(OH)_2$/C, MeOH, 32°C; (k) Ac_2O, pyridine; (l) NaOMe, MeOH; (m) BzCl, pyridine; (n) H_2, Pd/C MeOH; (o) i) LiHMDS, $[(BnO)_2P(O)]_2O$, -68°C to 0°C; ii) H_2, Pd/C MeOH; (p) 4-morpholine-N,N'-dicyclohexylcarboxamidinium uridine 5'-monophosphomorpholidate, tetrazole, pyridine; (q) NaOMe, MeOH

Figure 4-4. Original chemical synthesis of UDP-bacillosamine starting from commercially available galactosamine (8,11).

To achieve benzyl protection of the anomeric hydroxyl, the original protocol employed a two-step approach, involving the formation of an oxazoline followed by subsequent ring opening with benzyl alcohol to produce the product as a "pale brown solid" with an overall yield of 67%. Both the yield and purify of the product was improved though the use of a one-step protocol using BF_3 etherate as a Lewis acid, producing the desired product as a white crystals in 87% yield (**Figure 4-5**).

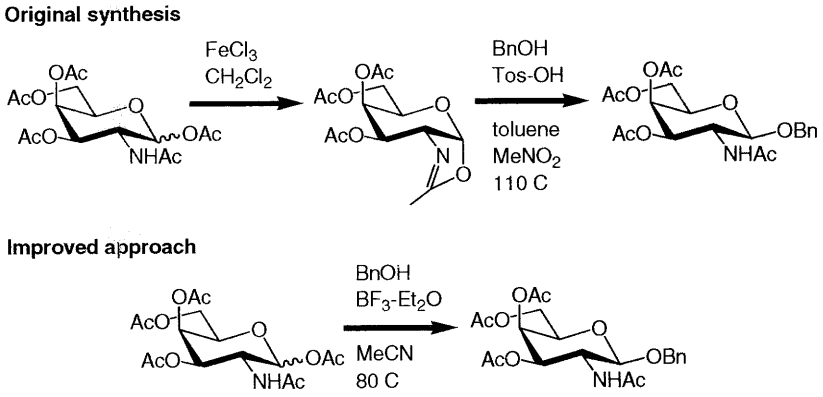


Figure 4-5. Improved approach to anomeric benzyl protection using BF_3 etherate.

For the purification of many chemical intermediates, the original protocol employed a chloroform/methanol solvent system for silica flash chromatography. However, due to the polar N-acetyl groups present in each intermediate, an ethylacetate/hexane/methanol system was found to provide superior separation amongst the compounds with and without the N-acetyl groups. The use of this solvent system led to greater purity of each chemical intermediate for NMR characterization.

Optimization of the original chemical synthesis led to an overall 12-fold improvement in yield and allowed publication-quality NMR spectra (11) to be obtained for many of the synthetic intermediates that were previously difficult to purify. Access to large amounts of UDP-bacillosamine was essential for the subsequent synthesis of undecaprenol-linked substrates that are necessary in the studies of enzymes in the Pgl pathway. Furthermore, this UDP-bacillosamine was also applied to the study of the related *N. gonorrhoeae* glycosylation pathway in our lab, when it was discovered that this pathway also uses UDP-bacillosamine as a key substrate.

4-2. Enzymatic synthesis of undecaprenyl diphosphate-linked compounds

Conveniently, each of the undecaprenol-linked intermediate can be synthesized in vitro using the glycosyltransferases from the Pgl pathway. PglC can be used to transfer bacillosamine phosphate from UDP-bacillosamine to undecaprenyl phosphate, creating the first undecaprenol-linked intermediate: undecaprenyl diphosphate-bacillosamine (8). PglA, J, and H can then sequentially add five N-acetylgalacosamine (GalNAcs) to this substrate from UDP-GalNAc, each with α-1,4 linkages (**Figure 4-1**) (12). Finally, PglI adds the final glucose to the third GalNAc in a β-1,3 linkage (12). Each enzyme was either recombinantly overexpressed in *E. coli* with a C-terminal His_6-tag and purified to homogeneity by Ni-NTA affinity chromatography, or they were used as a semi-pure membrane envelope fraction for ease of handling. These glycosyltransferases require no cofactor except for Mg^{2+} metal, and can function together efficiently in a one-pot 5-enzyme in vitro reaction to produce only the expected product (**Figure**

4-6) (12). The synthesis of any intermediates within the pathway can be carried out by simply excluding the appropriate glycosyltransferase and UDP-sugar donor.

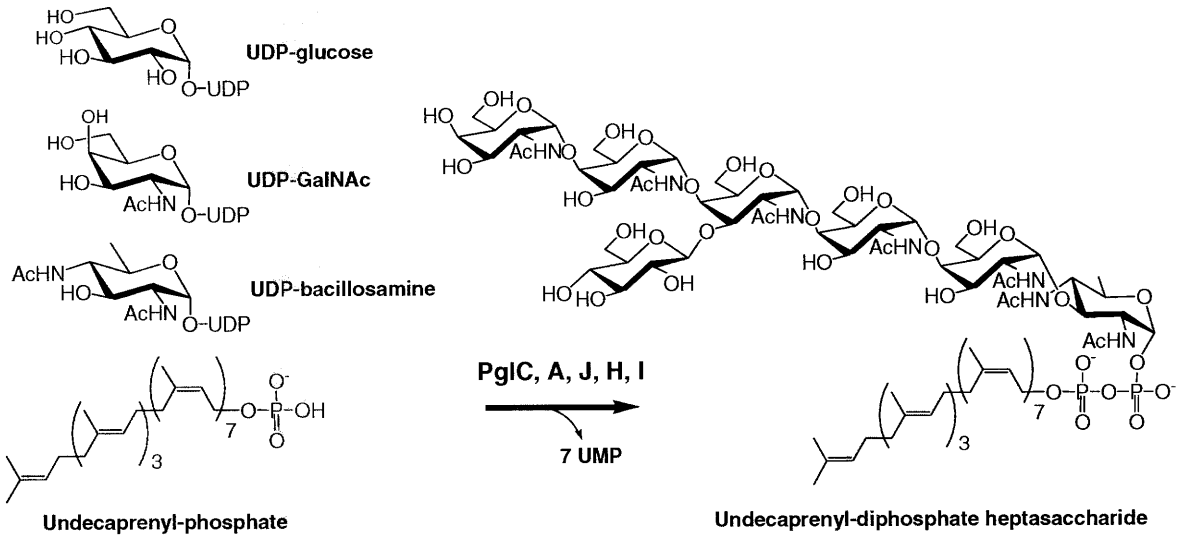


Figure 4-6. One pot enzymatic synthesis of undecaprenyl diphosphate-heptasaccharide, the native glycan donor in the *C. jejuni* N-linked glycosylation pathway.

The hydrophobic undecaprenyl-linked compounds can be isolated from the hydrophilic UDP-sugar starting materials by extraction using organic solvents (**Figure 4-7**). Upon completion of the reaction, usually lasting 2-3 hours, the reaction mixture is quenched with a chloroform/methanol solution, resulting in a phase separation between the aqueous and organic fractions. The water-soluble UDP-sugars remain in the aqueous fraction, while undecaprenol-linked compounds are extracted into to the organic fraction due to their extremely hydrophobic aliphatic component. Multiple extractions allow for nearly quantitative separation of those two reaction components. To purify the individual undecaprenyl-linked glycan intermediates, a semi-preparative normal-phase HPLC was used, which offers quick and effective baseline separation of intermediates that differ even by a single sugar.

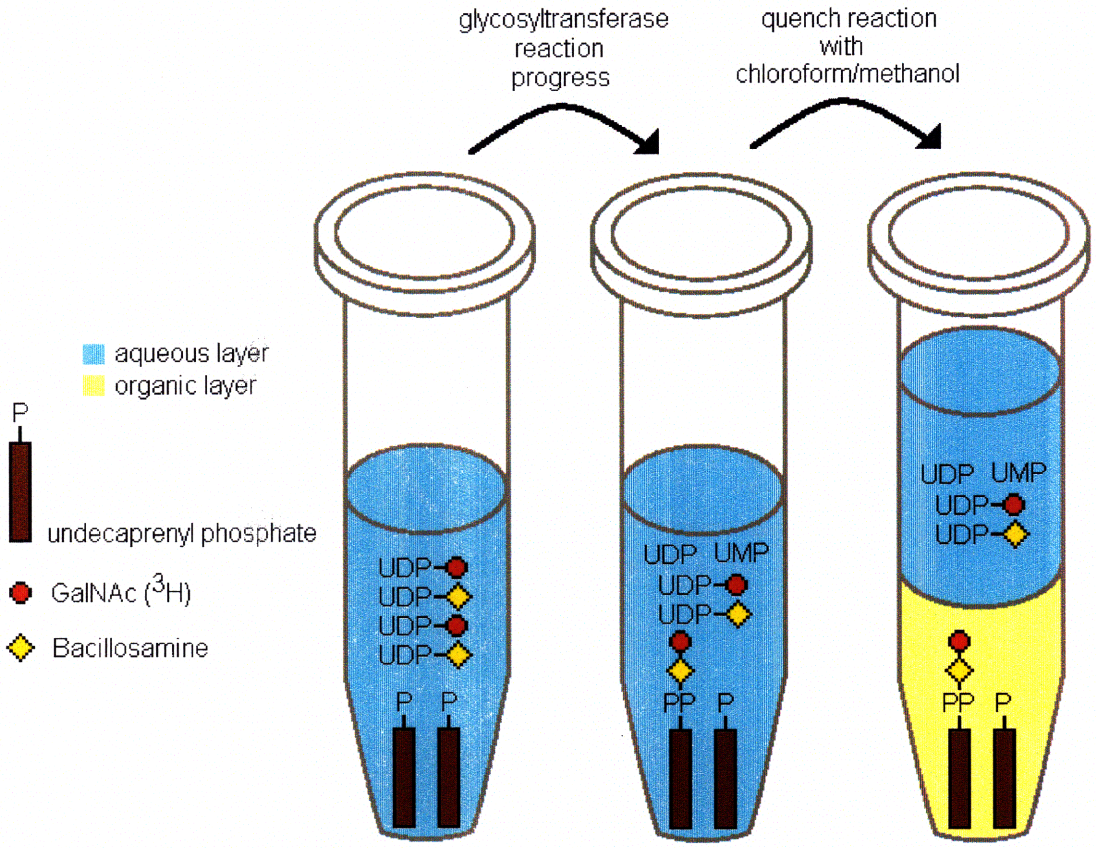


Figure 4-7. Separation of undecaprenol-linked products from excess starting material using organic solvents extractions, following a glycosyltransferase reaction.

An additional aspect of the enzymatic synthesis of undecaprenol-linked substrates was the incorporation of tritium radiolabels, which allowed for the detection and sensitive quantification of these substrates in enzymatic assays. Tritium radiolabels were typically introduced through the use of commercially available UDP-[C6-^{3}H]GalNAc, which can be spiked into the reaction mixture to create a known concentration of UDP-GalNAc of the desired specific activity, and any glycosyltransferase from the pathway that uses UDP-GalNAc. For glycosyltransferase assays, reactions were monitored based on the amount of water soluble UDP-[C6-^{3}H]GalNAc transferred to the organic soluble undecaprenyl diphosphate-carrier. For

oligosaccharyltransferase assays, reactions were monitored based on the amount of organic soluble undecaprenyl diphosphate-linked [^{3}H]glycan transferred to the water-soluble peptide acceptor. This ability to incorporate radiolabels is a major advantage, as non-radiolabeled substrates isolated from biological sources do not offer such a useful handle for quantification.

Progress that was later made in our laboratory allowed for a more convenient and quantitative biosynthesis of UDP-bacillosamine from UDP-GlcNAc using the three biosynthetic enzymes from the Pgl pathway: PglF, E, and D (**Figure 4-8**). Interestingly, it was found that these three enzymes can also be incorporated with the five glycosyltransferases to synthesize the full-length undecaprenyl diphosphate-heptasaccharide in a one-pot reaction starting with undecaprenyl phosphate and the commercially available UDP-sugar nucleotides (13). The specificity of PglC only allows bacillosamine to be transferred to the undecaprenol-carrier and does not appear to recognize the biosynthetic intermediates created by PglF or E. This new one-pot 8-enzyme synthesis offers the most convenient and efficient synthesis of undecaprenol-linked compounds, and represents an improvement over the previous chemical synthesis.

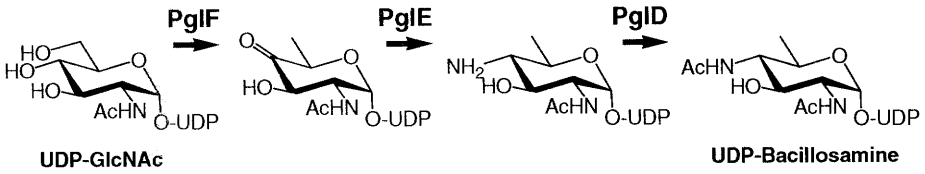


Figure 4-8. Biosynthesis of UDP-bacillosamine using Pgl F, E and D from the Pgl pathway.

4-3. Testing the binding of undecaprenyl diphosphate-linked substrates to lectins

The undecaprenol-linked substrates can find useful applications in any assay or study that requires a sensitive method for quantifying the attached glycan. Lectin binding was one such case where these compounds were used (14).

For example, a significant amount of in vivo research on the Pgl pathway was not carried out in the native *C. jejuni* host organism, but rather in a strain of *E. coli* transformed with the Pgl locus (5). This is mainly because *C. jejuni* is a pathogenic bacterium, while *E. coli* has already been established as a versatile organism for protein production and bioengineering. Although the glycosylation machinery appears to be functional within *E. coli*, it is important to determine whether the biosynthetic production of undecaprenol-linked substrates is the same as in the native organism. In order to make this comparison, our collaborators from the National Research Council of Canada designed a method of extracting and purifying the undecaprenol-linked substrates from cells and identifying the isolated samples by capillary electrophoresis mass spectrometry. Their extraction protocol is similar to our own, using chloroform/methanol solvents, but it also includes an additional enrichment step involving soybean agglutinin (SBA) lectins for affinity capture of GalNAc containing compounds (14). However there was no convenient method for them to evaluate the binding efficiency of the SBA lectin to determine whether it was reproducible and quantitative.

The tritium-labeled substrates produced from our synthesis were well suited to be used as standards for evaluating the binding of undecaprenyl-linked substrates to lectin. The radiolabel can easily allow detection of substrates down to the nanomolar concentration range. To evaluate

effectiveness of lectin binding, each of the GalNAc containing undecaprenyl-linked intermediate from the Pgl pathway was synthesized, purified, and then incubated with SBA resin. The supernatant was removed and the resin washed with buffer to remove unbound substrate. Finally the substrates were eluted from the SBA resin using a concentrated solution of GalNAc. The radioactivity from the unbound supernatant, the washes, the elution, and the resin can then be quantified by scintillation counting (14).

The first experiment involved a 1:1:1:1 mixture of undecaprenol-linked disaccharide, trisaccharide, hexasaccharide, and the full length native heptasaccharides. It was found that approximately 70% of the total substrate bound to the resin, while a significant amount remained in solution following the incubation (**Figure 4-9**). This incomplete binding does not appear to be due to insufficient resin, as doubling the amount of resin did not significantly improve binding.

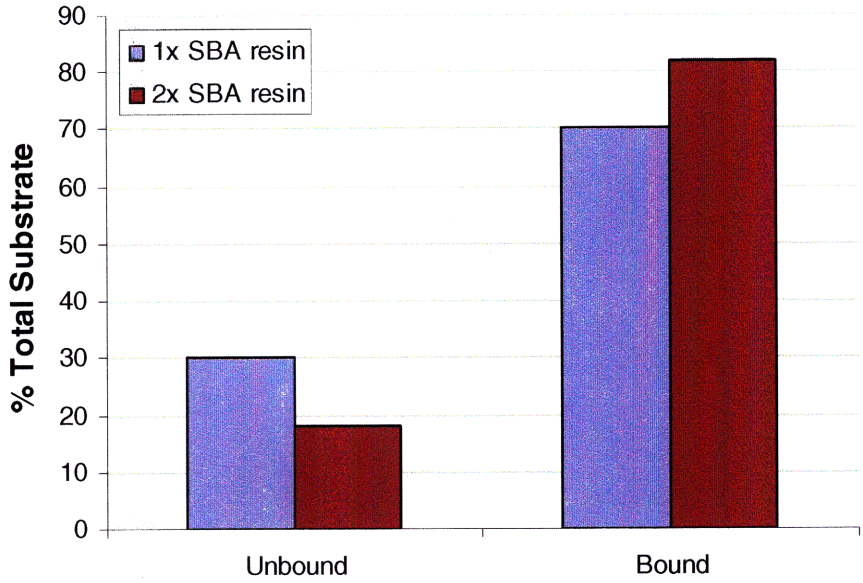


Figure 4-9. Evaluation of SBA-lectin binding of a 1:1:1:1 mixture of undecaprenyl diphosphate-linked disaccharide, trisaccharide, hexasaccharide, and the full length native heptasaccharides.

A second set of experiments were carried out to evaluate the binding of each substrate individually (**Figure 4-10**). A correlation was discovered between the number of GalNAc residues present in the substrate and the affinity of that substrate for the lectin resin. For the native heptasaccharide, approximately 85% of the substrate remained bound to the resin after washing. In contrast, for the disaccharide, containing a single GalNAc residue, only 40% of the substrate remained bound to the resin, while the rest either was released during the wash or never bound at all.

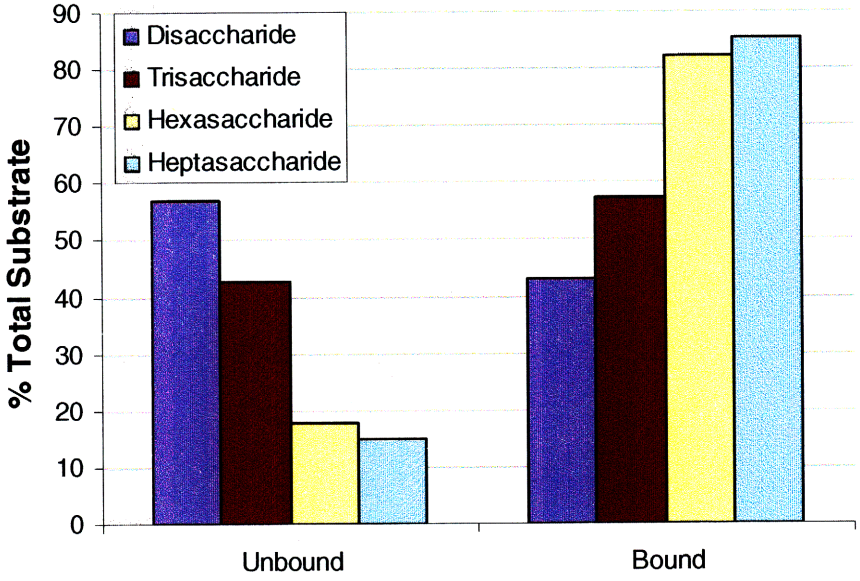


Figure 4-10. Evaluation of SBA-lectin binding of undecaprenyl diphosphate-linked disaccharide, trisaccharide, hexasaccharide, and heptasaccharides substrates individually.

Despite the incomplete binding, each of the undecaprenyl-linked substrates was confirmed to bind to the SBA resin. The results from these binding assays were remarkably reproducible, differing between experimental repeats by less than 5%. Therefore, when comparing the differences in substrate synthesis between *C. jejuni* and *E. coli*, affinity capture with SBA can be a viable method of isolation of GalNAc containing substrates. Since lectin binding is typically

known to be weak and cooperative, it was not surprising to observe that better binding was achieved by larger glycans that contained more GalNAc residues.

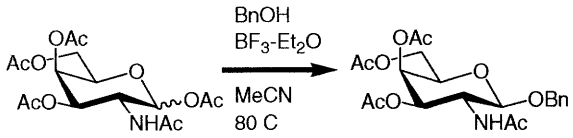
Conclusion.

In order to probe the enzymes involved in the newly discovered N-linked glycosylation pathway in *C. jejuni*, we developed an efficient chemoenzymatic in vitro synthesis for the unusual UDP-bacillosamine sugar as well as for each of the undecaprenol-linked substrate intermediates found in the pathway. Our synthetic strategy allowed us to access milligram-quantities of the desired substrates with high purity. Importantly, it allowed us to conveniently incorporate radiolabels into each compound as handles for accurate quantification in any in vitro assay, which is not possible with natively isolated substrates. Furthermore, it offered us the versatility of making unnatural analogues not available from native sources.

To demonstrate one practical use for these radiolabeled substrates, we synthesized four undecaprenol-linked glycans of varying size and evaluated their binding efficiency to SBA-lectin. Although each substrate was recognized by the lectin, there was a direct correlation between the number of GalNAc residues present in the substrate and the strength of the binding. Such an analysis could not have been made without access to pure, homogenous, and radiolabeled substrates. The next two chapters will describe further studies where these synthetic substrates were used to probe and characterize the enzymes of the Pgl pathway.

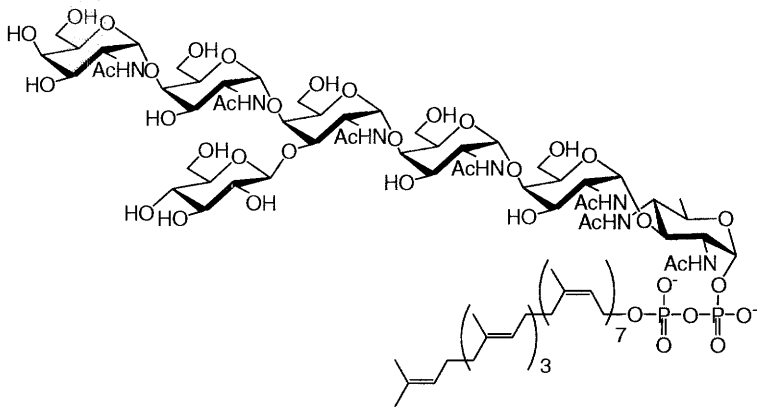
Experimental

Anomeric benzyl protection of peracetylated galacosamine.



Peracetylated galactosamine (17.25 g, 44 mmol) was dissolved in acetonitrile (75 mL, HPLC grade ≥99.9%). BF_3-diethyl etherate (300 µL, 2.4 mmol) and benzyl alcohol (10 mL, 100 mmol) was added to the mixture and heated to reflux at 80 °C for 3 hours. The acetonitrile solvent was removed under vacuum, and the precipitate was dissolved in 9:1 ethyl acetate / methanol for recrystallization, from which 14.4 g of highly pure benzyl glycoside product was isolated. The mother liquid was applied to flash chromatography using a 50:45:5 ethyl acetate / hexane / methanol solvent system, from which 2.8 g of pure benzyl glycoside product was isolated. Combined with the recrystallized product, an overall yield of 87% was obtained.

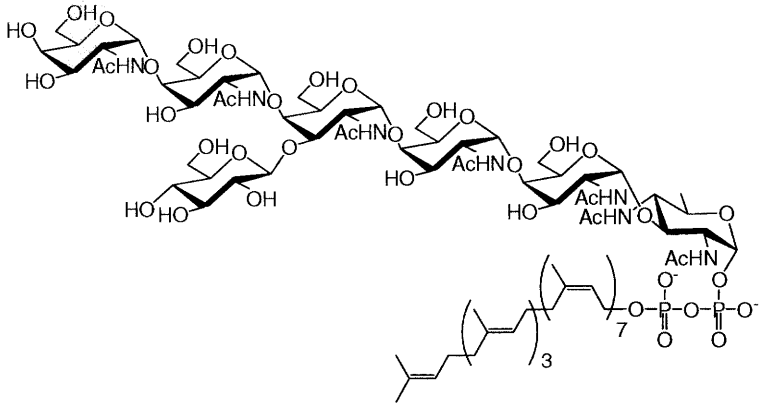
Synthesis of undecaprenyl diphosphate-heptasaccharide using UDP-bacillosamine.



Dried undecaprenyl phosphate (5 nmol) was resuspended using DMSO (3 µL) and Triton X-100 detergent (7 µL of 14.3% stock) with rigorous vortexing and sonication (water bath). To this

mixture was added UDP-bacillosamine (10 nmol), UDP-GalNAc (100 nmol), UDP-glucose (10 nmol), and brought to a total volume of 55 μL with final buffer concentrations of 50 mM tris-acetate, 10 mM $MgCl_2$, pH 8.5. Next, purified glycosyltransferases PglC, PglA, PglJ, and PglH were added (5 μL each, ~0.5 mg/mL), followed by PglI cell envelope fraction (25 μL). Enzyme preparation carried out as described in the literature (8,12). The reaction was thoroughly mixed and allowed to proceed for 120 min at room temperature. To isolate the undecaprenyl-linked products, the reaction mixture was quenched by the addition of PSUP (160 μL) and a 2:1 mixture of chloroform / methanol (800 μL). The tube was centrifuged briefly, and the organic (bottom) layer containing the desired products were removed and dried under reduced pressure.

Synthesis of undecaprenyl diphosphate-heptasaccharide using UDP-GlcNAc and bacillosamine biosynthesis enzymes.



Dried undecaprenyl phosphate (5 nmol) was resuspended using DMSO (3 μL) and Triton X-100 detergent (7 μL of 14.3% stock) with rigorous vortexing and sonication (water bath). To this mixture was added UDP-GlcNAc (100 nmol), UDP-GalNAc (100 nmol), UDP-glucose (20 nmol), NAD^+ (50 nmol), L-glutamate (2000 nmol), PLP (25 nmol), AcCoA (100 nmol) and brought to a total volume of 40 μL with final buffer concentrations of 50 mM tris-acetate, 10 mM $MgCl_2$, pH 8.5. Next, purified enzymes PglD, PglC, PglA, PglJ, and PglH were added (4

mM $MgCl_2$, pH 8.5. Next, purified enzymes PglD, PglC, PglA, PglJ, and PglH were added (4 μL each, ~0.5 mg/mL), followed by PglE (10 μL, ~0.5 mg/mL), PglF (20 μL, 0.15 mg/mL), and PglI cell envelope fraction (10 μL). Enzyme preparation was carried out as described in the literature (8,12,13). The reaction was thoroughly mixed and allowed to proceed for 7 hours at room temperature. To isolate the undecaprenyl-linked products, the reaction mixture was quenched by the addition of PSUP (160 μL) and a 2:1 mixture of chloroform / methanol (800 μL). The tube was centrifuged briefly, and the organic (bottom) layer containing the desired products were removed and dried under reduced pressure.

Synthesis of radiolabeled undecaprenyl-linked substrates.

To synthesize radiolabeled substrates, one of commercially available tritiated substrates (UDP-[C6-^{3}H]GalNAc, UDP-[C6-^{3}H]GlcNAc, UDP-[C6-^{3}H]Glucose, or [^{3}H]AcCoA) was substituted for the corresponding non-radiolableld component of the reaction. Typically the radiolabeled substrate was diluted with the corresponding non-radiolabeled to achieve the desired specific activity. Unless otherwise stated, radiolabel used in this thesis was incorporated as UDP-[C6-^{3}H]GalNAc.

Synthesis of truncated undecaprenyl-linked substrates.

To synthesize truncated undecaprenyl-linked substates, such as the undecaprenyl diphosphate-disaccharide, trisaccharide, and hexasaccharide, the appropriate glycosyltransferase and UDP-glucose was neglected from the reaction, as well as a reduction in UDP-GalNAc concentration depending on stoichiometry.

Purification of undecaprenyl-linked substrates.

Undecaprenyl-linked substrates isolated from the reaction mixture by organic extraction was dried under reduced pressure and resuspended in 2:1 mixture of chloroform / mechanol. This mixture was separated on a normal-phase Varian Microsorb HPLC column using the following gradients at 1 mL/min flow rate: 0-3 min at 0% B; 3-5 min at 0-20% B; 5-35 min at 20-30% B; 35-60 min at 30-45% B; 65-70 min at 100% B; where A = 4:1 chloroform / methanol, and B = 10:10:3 chloroform / methanol / 2 M ammonium acetate.

Undecaprenyl-linked substrate lectin binding assays.

Dried undecaprenyl diphosphate-oligosaccharide[^{3}H-GalNAc] (0.13 nmol) was resuspended using a 9:1 mixture of ethanol / methanol (40 µL) with rigorous vortexing and sonication. It was diluted with 360 µL of buffer A [10 mM HEPES (*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic sodium salt), 3.4 mM EDTA (ethylenediaminetetracetic acid), 150 mM NaCl, pH 7.4], followed by the addition of SBA resin (20 µL) pre-equilibrated with buffer A. This mixture was left to rotate at 4 °C for 2 hours. Supernatant was then removed and the resin washed with buffer A (400 µL). The resin was then incubated for 3 x 15 min with 400 / 200 / 200 µL of 200 mM galactose solution at each incubation. The unsuspended residue, the supernatant and wash, the elution, and the resin were individually subjected to scintillation counting. Each experiment was carried out in duplicate.

References

1. Young NM, Brisson JR, Kelly J, Watson DC, Tessier L, Lanthier PH, Jarrell HC, Cadotte N, St Michael F, Aberg E, Szymanski CM (2002) Structure of the N-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, *Campylobacter jejuni*. *J. Biol. Chem.* **277**: 42530-9.

2. Szymanski CM, Michael FS, Jarrell HC, Li J, Gilbert M, Larocque S, Vinogradov E, Brisson JR (2003) Detection of conserved N-linked glycans and phase-variable lipooligosaccharides and capsules from *Campylobacter* cells by mass spectrometry and high resolution magic angle spinning NMR spectroscopy. *J. Biol. Chem.* **278**: 24509-20.

3. Szymanski CM, Logan SM, Linton D, Wren BW (2003) *Campylobacter*--a tale of two protein glycosylation systems. *Trends. Microbiol.* **11**: 233-8.

4. Linton D, Dorrell N, Hitchen PG, Amber S, Karlyshev AV, Morris HR, Dell A, Valvano MA, Aebi M, Wren BW (2005) Functional analysis of the *Campylobacter jejuni* N-linked protein glycosylation pathway. *Mol. Microbiol.* **55**: 1695-703.

5. Wacker M, Linton D, Hitchen PG, Nita-Lazar M, Haslam SM, North SJ, Panico M, Morris HR, Dell A, Wren BW, Aebi M (2002) N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science* **298**: 1790-3.

6. Weerapana E, Imperiali B (2006) Asparagine-linked protein glycosylation: From eukaryotic to prokaryotic systems. *Glycobiology* **16**: 91-101.

7. Kelly J, Jarrell H, Millar L, Tessier L, Fiori LM, Lau PC, Allan B, Szymanski CM (2006) Biosynthesis of the N-linked glycan in *Campylobacter jejuni* and addition onto protein through block transfer. *J. Bacteriol.* **188**: 2427-34.

8. Glover KJ, Weerapana E, Chen MM, Imperiali B (2006) Direct biochemical evidence for the utilization of UDP-bacillosamine by PglC, an essential glycosyl-1-phosphate transferase in the *Campylobacter jejuni* N-linked glycosylation pathway. *Biochemistry* **45**: 5343-50.

9. Sharon N (2007) Celebrating the golden anniversary of the discovery of bacillosamine, the diamino sugar of a Bacillus. *Glycobiology* **17**: 1150-5.

10. Aas FE, Vik A, Vedde J, Koomey M, Egge-Jacobsen W (2007) Neisseria gonorrhoeae O-linked pilin glycosylation: functional analyses define both the biosynthetic pathway and glycan structure. *Mol Microbiol* **65**: 607-24.

11. Weerapana E, Glover KJ, Chen MM, Imperiali B (2005) Investigating bacterial N-linked glycosylation: synthesis and glycosyl acceptor activity of the undecaprenyl pyrophosphate-linked bacillosamine. *J. Am. Chem. Soc.* **127**: 13766-7.

12. Glover KJ, Weerapana E, Imperiali B (2005) *In vitro* assembly of the undecaprenylpyrophosphate-linked heptasaccharide for prokaryotic N-linked glycosylation. *Proc. Natl. Acad. Sci. U.S.A.* **102**: 14255-9.

13. Olivier NB, Chen MM, Behr JR, Imperiali B (2006) *In Vitro* Biosynthesis of UDP-N,N'-Diacetylbacillosamine by Enzymes of the *Campylobacter jejuni* General Protein Glycosylation System. *Biochemistry* **45**: 13659-69.

14. Reid CW, Stupak J, Chen MM, Imperiali B, Li J, Szymanski CM (2008) Affinity-capture tandem mass spectrometric characterization of polyprenyl-linked oligosaccharides: tool to study protein N-glycosylation pathways. *Anal Chem* **80**: 5468-75.

CHAPTER 5

OLIGOSACCHARYLTRANSFERASE PGLB ACCEPTOR SPECIFICITY

Acknowledgements

The UDP-bacillosamine used in this study was provided by Dr. Nelson Olivier.

The $Im7_{DQNAT}$ mutant was cloned, expressed, and assayed by Dr. Jebrel Glover.

A significant portion of this chapter was published in:

Chen MM, Glover KJ, Imperiali B (2007) From Peptide to Protein: Comparative Analysis of the Substrate Specificity of N-Linked Glycosylation in C. jejuni. *Biochemistry* **46**: 5579-85.

In the eukaryotic kingdom, N-linked glycosylation takes place on the asparagine side chain of the acceptor protein within the consensus sequence Asn-X-Ser/Thr, where X can be any amino acid except for proline (1-3). Although this sequence is necessary for glycosylation to occur, it is not sufficient, as many potential glycosylation sites have been known to be either unglycosylated or inefficiently processed (4). Factors that influence glycosylation site occupancy and efficiency include the choice of serine or threonine hydroxy-amino acid, the local sequence and secondary structure, the amino acid in the X-position, and the proximity of the glycosylation site to the C-terminus of the protein (5).

Recently, *Campylobacter jejuni* became the first and, thus far, only bacterium discovered to contain a general N-linked protein glycosylation system (6,7). Although the specific substrates are different, this protein glycosylation (Pgl) pathway appears to be very analogous to the eukaryotic pathway. Central to this pathway is the oligosaccharyltransferase (OT) PglB, an 82 kDa integral membrane protein with significant sequence homology to the Stt3p subunit of the eukaryotic OT, which is believed to contain the catalytic domain (8,9). PglB catalyzes the transfer of the preassembled heptasaccharide from the undecaprenyl diphosphate-carrier to the asparagine side chain of the polypeptide acceptor. We have previously demonstrated the OT activity of PglB in vitro using a synthetic radiolabeled disaccharide glycosyl donor and a short peptide acceptor (10). However, little is known about the acceptor sequence preferences for this prokaryotic OT, which is crucial for understanding the factors that govern glycan site occupancy and how it differs in comparison to the better known eukaryotic system. Furthermore, information regarding the binding preferences of PglB can aid in the design of tight binding inhibitor for mechanistic studies.

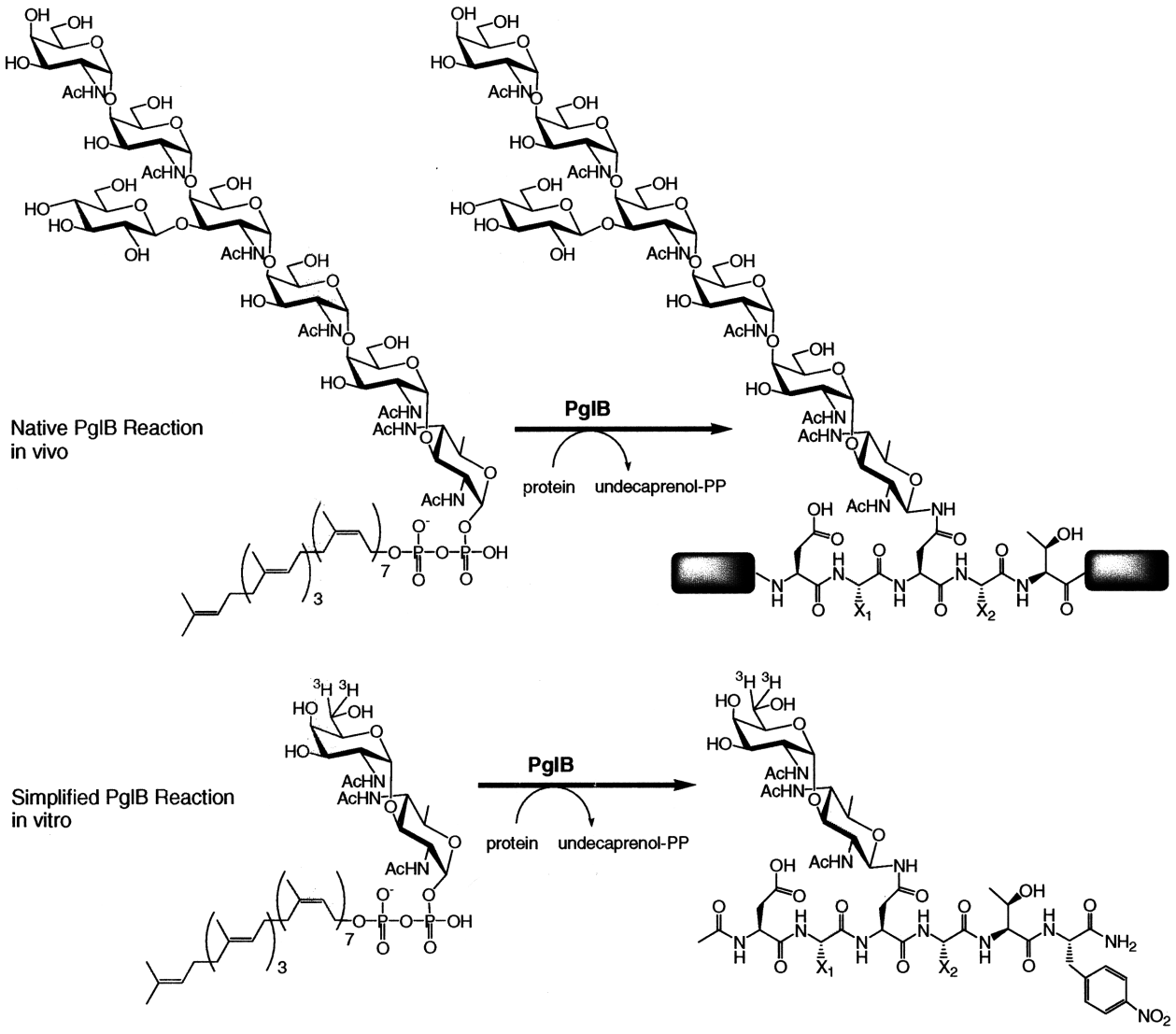


Figure 5-1. Comparison of native PglB reaction in vivo and the tritium labeled simplified reaction for assays in vitro. The rectangles represent the full length protein.

5-1. Design and synthesis of PglB peptide acceptors

In order to assay PglB in vitro, both the glycan and the protein acceptor were slightly modified. Instead of the full-size native heptasaccharide, we used a truncated tritium-labeled disaccharide (undecaprenyl diphosphate-bacillosamine-α-1,3-GalNAc) (**Figure 5-1**). The radiolabel provides the signal sensitivity needed to carry out the assay in the low micromolar concentration range,

which would not be possible with native substrates. The smaller disaccharide is easier to synthesize (11), has better solubility properties, and has already been demonstrated to be well accepted by PglB (10). Although N-linked glycans are only found attached to proteins in vivo, both the eukaryotic OT and PglB have been demonstrated to glycosylate small flexible peptides in vitro. Peptides make ideal acceptors for in vitro assays because they can be easily synthesized by solid phase, they can be made to include unnatural amino acids to probe the binding of the enzyme, and they contain virtually any sequence without concern for disrupting the protein fold.

In this study we designed and synthesized a library of peptide acceptor substrates for PglB. Using a sensitive radioactivity-based in vitro assay, we proceeded to define the minimal glycosylation consensus sequence recognized by PglB and the amino acid preference at each residue location. Our peptide template was based on a known glycosylation site ^{88}DFNVS^{92} from the *C. jejuni* glycoprotein PEB3 (12,13). A *para*-nitrophenylalanine (*p*NF) (**Figure 5-1**) was appended to the C-terminus of each peptide to facilitate accurate peptide concentration determination by UV/Vis spectroscopy. All peptides were capped at both termini to simulate an interior peptide sequence, because this has been previously shown to improve binding for the eukaryotic OT (2). Each peptide only differed by a single amino acid, so that the contribution of each residue within the glycosylation sequence could be systematically evaluated (**Figure 5-2**).

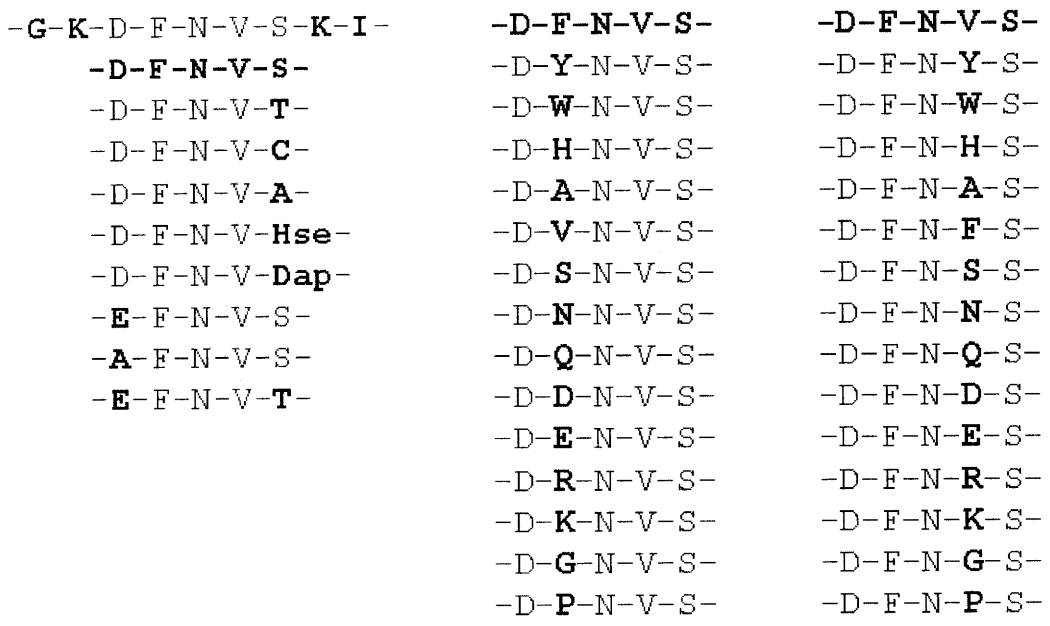


Figure 5-2. Sequence of peptides synthesized for evaluation of PglB. Each peptide contained a C-terminus *para*-nitrophenylalanine for concentration determination. Hse = homoserine, side chain extended from serine by one methylene group. Dap = 2,3-diaminopropionic acid, side chain truncated from lysine by two methylene groups.

5-2. Evaluation of PglB glycosylation sequence preferences

The consensus N-linked glycosylation sequence in *C. jejuni* shared both interesting similarities and differences with the eukaryotic counterpart. One major difference in *C. jejuni* glycosylation was the additional requirement of an acidic residue (aspartate or glutamate) at the -2 position. Although under high substrate concentrations we observed glycosylation of the Asn-Leu-Thr peptide in vitro (10), the acidic residue appeared to be necessary for efficient glycosylation to take place (**Figure 5-3**). Although aspartate and glutamate are equally common at the -2 position of N-linked glycosylation site from *C. jejuni* glycoproteins, we found that PglB exhibited a distinct preference for aspartate at that position. Between the two model peptides Ac-EVNAT-

(*p*NF)-NH_2 and Ac-DVNAT-(*p*NF)-NH_2, having glutamate as the acidic residue at position -2 increased the $K_{m(app)}$ by six-fold, while decreasing the $V_{max(app)}$ by five-fold (**Table 5-1**).

Eukaryotic – Asn – X – Ser/Thr –

C. jejuni – Asp/Glu – X_1 – Asn – X_2 – Ser/Thr –

–2 –1 0 +1 +2

where X ≠ Pro

Figure 5-3. Minimum glycosylation consensus sequence of the eukaryotic and *C. jejuni* OT.

Peptides containing threonine, as the hydroxyamino acid at position +2, exhibited a three-fold lower $K_{m(app)}$ in comparison to serine (**Table 5-1**). This preference for threonine is shared by the eukaryotic OT, wherein substitution of serine for threonine can dramatically improve the glycosylation efficiency of a poorly-glycosylated site in vivo (14,15). The effects of the substitutions appeared to be additive. When both unfavorable amino acids were present within the same peptide Ac-EFNVS-(*p*NF)-NH_2, glycosylation was negligible under the experimental conditions employed. This was initially surprising since the *C. jejuni* glycoprotein HisJ is known to be glycosylated at the site ^{27}ESNAS^{31}, although somewhat less efficiently than other *C. jejuni* glycoproteins (12).

Peptide conformation appeared to be an important factor for substrate recognition, as was shown for the eukaryotic OT (16). The unnatural peptides containing structural analogs cysteine, homoserine, or diaminopropionic acid at the +2 position were not found to be glycosylated by PglB to a noticeable degree under the assay conditions employed. Furthermore, these three

peptides did not exert any inhibitory effect on the glycosylation of other acceptor peptides, suggesting that they are not well recognized by PglB. Our interpretation of the finding is that the poor glycosylation activity is due to the fact that these peptides are less capable of forming the Asx-turn, a motif suggested to be recognized by the eukaryotic OT (17-19).

Table 5-1. Kinetic parameters of PglB peptide substrates. Unnatural amino acids: *p*NF = *para*-nitrophenylalanine, Hse = homoserine, Dap = 2,3-diaminopropionic acid.

Peptide	$V_{max(app)}$ (nM/min)	$K_{m(app)}$ (μM)	$V_{max(app)}/K_{m(app)}$ (min^{-1} x 10^3)
Ac-DFNVA-(*p*NF)-NH_2		no activity	
Ac-DFNVT-(*p*NF)-NH_2	34.6 ± 0.5	1.22 ± 0.07	28.4 ± 1.80
Ac-DFNVS-(*p*NF)-NH_2	34.6 ± 1.2	3.83 ± 0.55	9.03 ± 1.45
Ac-AFNVT-(*p*NF)-NH_2		no activity	
Ac-EFNVT-(*p*NF)-NH_2	6.9 ± 0.3	23.3 ± 2.98	0.30 ± 0.04
Ac-EFNVS-(*p*NF)-NH_2		no activity	
Ac-GKDFNVSKI-(*p*NF)-NH_2	11.3 ± 0.3	1.17 ± 0.14	9.67 ± 1.26
Ac-DFNVC-(*p*NF)-NH_2		no activity	
Ac-DFNV-(Hse)-(*p*NF)-NH_2		no activity	
Ac-DFNV-(Dap)-(*p*NF)-NH_2		no activity	

In the same manner, we evaluated the preferences of PglB for the amino acids in the two X-positions (**Figure 5-4**). For the X_1-position, PglB revealed a clear preference for the amido functionalities of asparagine and glutamine, as well as for large hydrophobic groups, whereas charged amino acids were clearly disfavored at that position. For the X_2-position, PglB exhibited trends similar to the eukaryotic OT (20), in which the positively charged groups of lysine and arginine, along with alanine and serine were favored, while large hydrophobic groups were disfavored. The differences within this set of peptides mainly comes from the differences in $K_{m(app)}$, while the relative $V_{max(app)}$ values are similar and do not follow the same trend. Like the

eukaryotic OT, proline was not accepted at either of the two X-positions. It is interesting that PglB has such contrasting preference for the two residues directly adjacent to the glycosylation site. Within the peptide library presented, the DQNAT sequon was found to be the optimal peptidyl substrate for PglB.

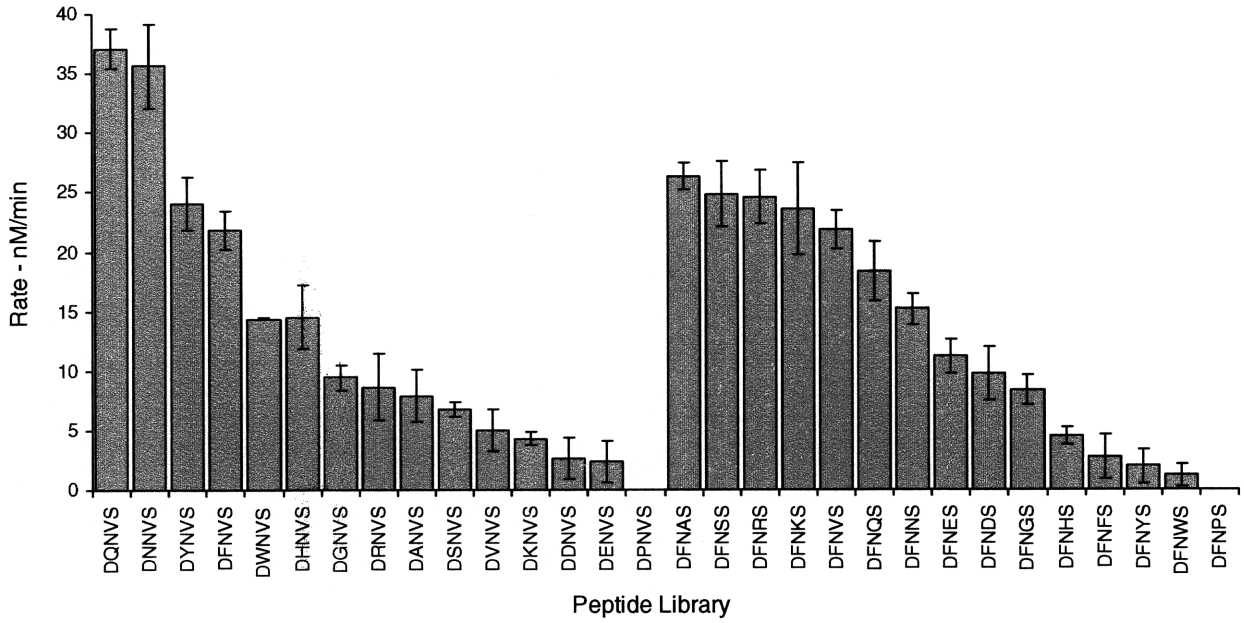


Figure 5-4. Comparison of PglB initial rates with peptides substrates varying at the X_1 and X_2 positions. Peptides were assayed at 10 μM concentration, close to the average peptide $K_{m(app)}$.

The extended PEB3 peptide Ac-GKDFNVSKI-(*p*NF)-NH_2 was found to have a lower $K_{m(app)}$ than its corresponding hexapeptide counterpart, indicating the presence of additional binding determinants in the substrate beyond the minimum consensus sequence (5). However, direct comparisons are complicated by the *p*NF residue being in a different a position from the rest of the peptides within the library.

5-3. Preparation of full-length folded protein as acceptors for PglB

In eukaryotic systems, N-linked glycosylation takes place as a co-translational process, in which the glycan is transferred to the nascent protein acceptor while the protein is being synthesized by the ribosome and threaded through the ER membrane by the translocation machinery (21). The eukaryotic OT is not capable of glycosylating folded proteins, and has a clearly demonstrated requirement for unstructured and flexible acceptors (22). In *C. jejuni*, protein synthesis takes place in the cytoplasm, while N-linked glycosylation takes place in the periplasm (21). It is unclear whether the glycan is transferred in a homologous manner to the eukaryotic system as the protein is translocated into the periplasm, or whether it occurs on fully folded protein acceptors already present in the periplasm.

To determine whether PglB has the ability to glycosylate a full-length folded protein, we engineered a Asp/Glu-X_1-Asn-X_2-Ser/Thr glycosylation consensus sequence into a well-defined loop region within Im7 (23-25), a small bacterial non-*C. jejuni* protein which is not glycosylated (**Figure 5-5**). Loops are the most common type of secondary structure where N-linked glycosylation is found, and are expected to be more tolerant of the mutations needed to create a PglB glycosylation sequence. Point mutations N26D and A28N resulted in the formation of a glycosylation consensus sequence ^{26}DVNAT^{30} within the loop region between the first and second helices of Im7. The overall structure of the Im7 mutant was compared with the wild-type using circular dichroism spectroscopy (CD) in order to verify that introduction of the glycosylation sequence did not significantly change the overall fold of the protein (**Figure 5-6**).

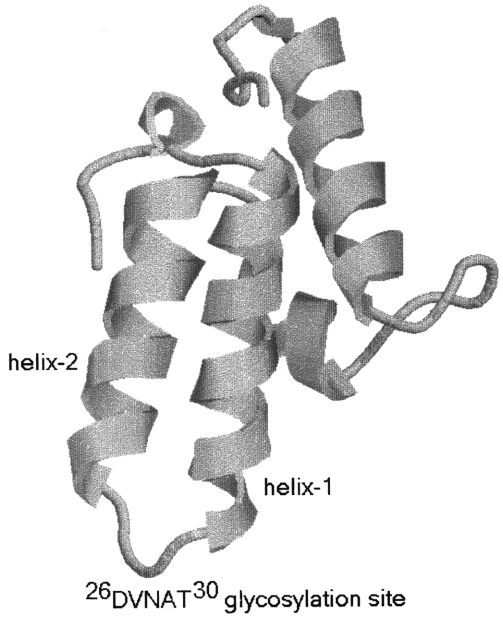


Figure 5-5. Native Im7 structure, indicating the chosen glycosylation site between helix-I and helix-II (26), pdb code 1CEI.

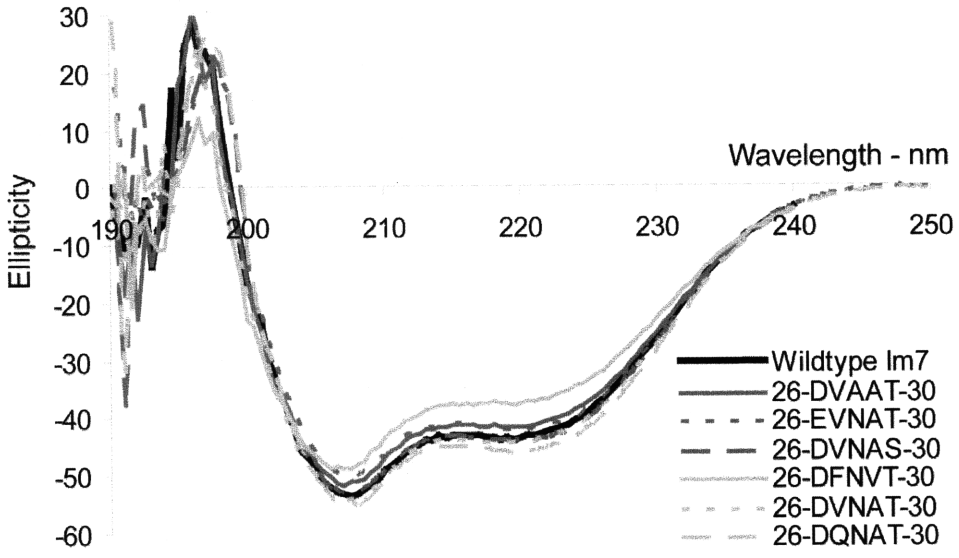


Figure 5-6. Comparison of Im7 mutants to wildtype Im7 secondary structures by circular dichroism spectroscopy.

5-4. PglB glycosylation of full length folded proteins

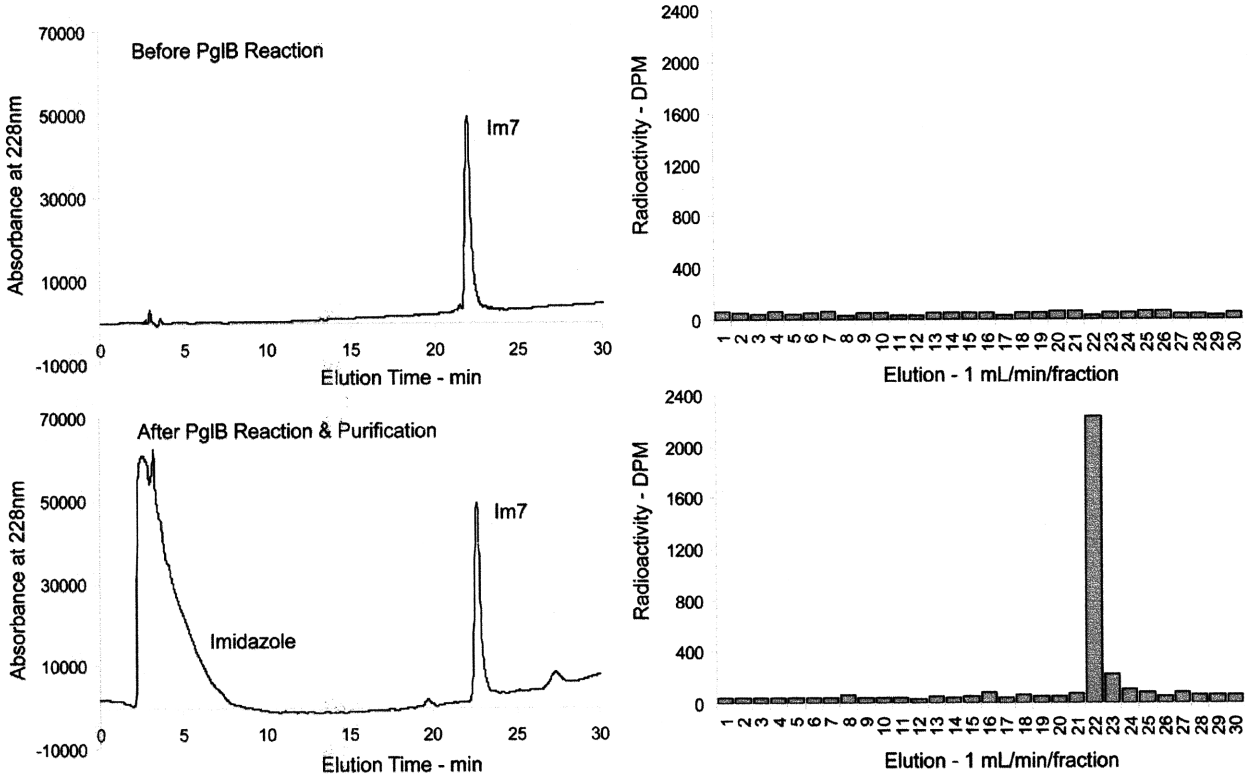


Figure 5-7. Reverse-phase HPLC traces of Im7 mutant ^{26}DVNAT^{30} before and after glycosylation by PglB. Elution of protein was monitored by absorbance at 228 nm, and collected in 1 mL fractions for scintillation counting. The extent of glycosylation was quantified by the amount of radiolabeled disaccharide co-eluting with the protein acceptor.

The $Im7_{DVNAT}$ mutant was incubated with PglB and the radiolabeled glycosyl donor for 3 hours. When the sample was subjected to a reverse-phase HPLC analysis, a significant amount of radioactivity was found to co-elute with a 228 nm absorption peak corresponding to the Im7 protein (**Figure 5-7**). No appreciable radioactivity was found to be associated with the wild-type $Im7_{NVAAT}$ and the non-glycosylatable $Im7_{DVAAT}$ mutant, missing the asparagine glycosylation site. As a negative control, radioactivity was not found to be associated with the Im7 mutant when PglB was not present, or when the eukaryotic OT was used with its native glycan donor

substrate (data not shown). Mass spectral analysis confirmed that the disaccharide was indeed covalently linked to the $Im7_{DVNAT}$ mutant (**Figure 5-8**). It appears that a major difference between prokaryotic and eukaryotic N-linked glycosylation is that PglB possesses the ability to glycosylate already folded proteins independent of the translocation machinery, whereas the eukaryotic OT requires the acceptor to be flexible and unconstrained by tertiary structure.

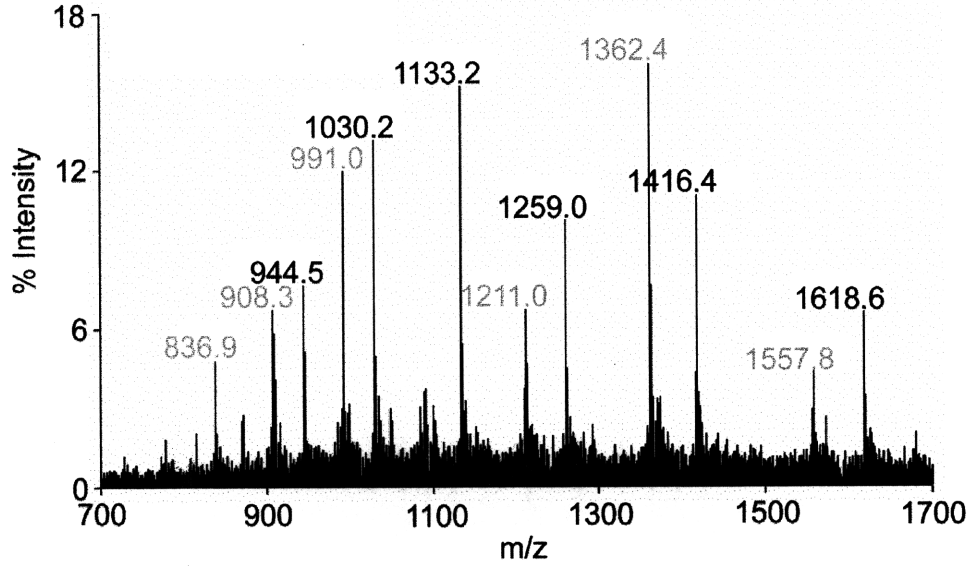


Figure 5-8. Electrospray ionization mass spectral analysis of glycosylated Im7 mutant ^{26}DVNAT^{30} following PglB glycosylation reaction. Masses in grey correspond to remaining unglycosylated Im7, while masses in black correspond to Im7 plus the disaccharide.

To determine whether the preferences and trends observed in the peptide acceptor study are relevant in the context of a fully folding protein, a small set of Im7 mutants varying in the glycosylation sequence was created to compare their ability as glycosylation acceptors. Like the previous mutant, each new Im7 was checked by CD to ensure that the mutations did not affect the structure of the protein. When assayed with PglB, the relative extents of glycosylation were

consistent with the glycosylation efficiencies of the corresponding peptides. However, the differences amongst the full-length proteins were less dramatic (**Table 5-2**). Our interpretation of this finding is that, similar to the eukaryotic OT, the acceptor binding of PglB is influenced by residues beyond the five amino acid consensus sequence, and that local structural conformation near the glycosylation site diminishes the effects of the glycosylation sequence. Nevertheless, $Im7_{DQNAT}$ mutant with the optimal glycosylation sequence as determined by the peptide study was the most efficiently glycosylated by PglB, indicating that the overall residue preferences of PglB are the same regardless of whether the substrate is a structured protein or a flexible peptide.

Table 5-2. Kinetic parameters of PglB peptide substrates in comparison to the glycosylation of corresponding Im7 mutants. The extent of glycosylation corresponds to the percentage of (limiting) radiolabeled glycosyl donor consumed in the 3-hour reaction.

Peptide	$V_{max(app)}$ (nM/min)	$K_{m(app)}$ (μM)	$V_{max(app)}/K_{m(app)}$ (min^{-1} x 10^3)	Im7 Mutant	Extent of Glycosylation
Ac-NVAAT-(*p*NF)-NH_2		no activity		**Native**	0 ± 0.0
Ac-DVAAT-(*p*NF)-NH_2		no activity		**^{26}DVNAT^{30}**	0 ± 0.0
Ac-EVNAT-(*p*NF)-NH_2	50.6 ± 1.6	22.5 ± 2.56	2.3 ± 0.29	**^{26}EVNAT^{30}**	22 ± 6.3
Ac-DVNAS-(*p*NF)-NH_2	31.0 ± 1.8	3.33 ± 0.43	9.3 ± 1.47	**^{26}DVNAS^{30}**	35 ± 5.0
Ac-DFNVT-(*p*NF)-NH_2	34.6 ± 0.5	1.22 ± 0.07	28.4 ± 1.80	**^{26}DFNVT^{30}**	43 ± 1.6
Ac-DVNAT-(*p*NF)-NH_2	32.1 ± 1.1	1.06 ± 0.05	30.3 ± 1.96	**^{26}DVNAT^{30}**	53 ± 3.2
Ac-DQNAT-(*p*NF)-NH_2	34.8 ± 1.2	0.80 ± 0.11	43.4 ± 6.73	**^{26}DQNAT^{30}**	60 ± 4.5

Conclusion

To determine the acceptor specificity of the *C. jejuni* oligosaccharyltransferase PglB, a library of peptidyl substrates was synthesized and assayed in vitro. Each peptide differed by a single amino acid so that the contributions of each residue position could be systematically evaluated. The minimum glycosylation consensus sequence was determined to be Asp/Glu-X_1-Asn-X_2-

Ser/Thr, where X can be any amino acid except for proline. Interestingly, although our limited list of known *C. jejuni* glycoproteins do not display a preference for residues Asp or Glu at position -2, and residues Ser and Thr at position +2, our sensitive assays revealed a distinct preference for Asp and Thr at those positions, respectively. Furthermore, trends can be observed for the two middle positions, where amido and aromatic residues are preferred at X_1, while small and positively charged residues are preferred at X_2.

To determine whether PglB has the ability to glycosylate fully folded proteins independent of the translocation machinery, the minimum glycosylation consensus sequence was engineered into a non-*C. jejuni* protein which is not naturally glycosylated. When PglB was incubated with this fully folded protein in vitro, glycosylation was confirmed. This ability is unique to PglB, as the eukaryotic OT can only glycosylate unstructured and flexible acceptors. The efficiency of glycosylation was improved when the glycosylation site was engineered to contain the optimal sequence as determined by the peptide study, demonstrating that the trends and preferences observed in the peptide study of PglB are relevant in the context of a folded protein.

Experimental

Peptide Synthesis.

All peptides were synthesized by automated peptide synthesis on an ABI 431A peptide synthesizer (Applied Biosystems) using standard Fmoc-based peptide synthesis conditions on PAL-PEG-PS resin. Each peptide was acetylated at the N-terminus, and includes a *para*-nitrophenylalanine (*p*NF) at the C-terminus. The peptides were cleaved from the resin using a trifluoroacetic acid cocktail containing 2.5% water and 2.5% triisopropylsilane, purified to ≥95% purity by preparative reverse-phase HPLC using a standard water:acetonitrile gradient, and quantified using the UV absorbance of the *p*NF amino acid at 280 nm (ε = 12,500 $M^{-1}cm^{-1}$).

Cloning and mutagenesis of Im7.

A pTrc (Im7) vector was provided by Professor Sheena Radford from the University of Leeds, encoding the native Im7 protein with an N-terminal hexa-histidine tag described elsewhere (23-25). Introduction of point mutations in Im7 was performed using the Quikchange mutagenesis kit (Stratagene). All mutants were sequenced to ensure that the gene contained the desired change.

Expression of Im7 Mutants.

Starting from a 5 mL overnight culture, *E. coli* strains expressing Im7 variants were grown at 37 °C in LB broth to an OD_{600} of 0.6-0.8. At that point, the temperature was reduced to 16 °C and protein production was induced by the addition of isopropyl-β-D-thiogalactopyranoside (1 mM). After 24 h, the cells were harvested by centrifugation (7,500 x *g*) for 30 min, washed once with a

0.9% NaCl solution, centrifuged again (7,500 x *g*) for 30 min, and the cell pellet was frozen at -80°C until needed.

Purification of Im7 Mutants.

All steps were performed at 4°C. Cell pellets of *E. coli* strains expressing Im7 variants were thawed and resuspended in 5% of the original culture volume in buffer L [50 mM Tris-acetate, 5 mM imidazole, pH 8]. The cells were lysed by sonication, followed by centrifugation (142,400 x g) for 1 h to remove cellular debris and membrane proteins. The supernatant was slowly applied to a column containing Ni-NTA agarose equilibrated with buffer L. After washing with 5 column volumes of buffer L, the purified protein was eluted with buffer E [50 mM Tris-acetate, 250 mM imidazole, pH 8]. Fractions containing a significant amount of desired protein were combined, and applied to a Superdex75 gel filtration column (Amersham Biosciences) eluting at 1 mL/min with buffer F [50 mM Tris-acetate, pH 8]. Final samples used were confirmed by ESI-MS to be within 1 Da of expected mass, and ≥95% pure by Coomassie staining of SDS-PAGE.

Synthesis of Radioactive Disaccharide Donor.

Radiolabeled [^{3}H]GalNAc-(α1,4)-Bac-(α1)-PP-undecaprenyl was synthesized following a protocol previously described (10). Briefly, chemically synthesized UDP-Bac and undecaprenyl phosphate and commercially available [^{3}H]UDP-GalNAc were incubated with the *C. jejuni* glycosyltransferases PglC and PglA. The radiolabeled disaccharide product was extracted into a 2:1 mixture of chloroform/methanol, aliquoted, and dried into 1.5 mL tubes, each with

approximately 0.3 nmol of disaccharide product (specific activity = 356,000 DPM/nmol, ~100,000 DPM/tube).

Glycosylation of Peptide Substrates Using PglB.

To a tube containing 0.3 nmol of dried radiolabeled [^{3}H]GalNAc-Bac-PP-undecaprenyl was added 10 µL of DMSO. Following vigorous vortexing and sonication (water bath) to resuspend the isoprene-based substrate, 100 µL of 2X assay buffer B [280 mM sucrose, 2.4% Triton X-100 (v/v), 280 mM Hepes pH 7.5], 2 µL of 1 M $MnCl_2$, and 6 µL of PglB membrane fraction containing approximately 50 ng of enzyme were added, and the volume increased to 190 µL with water. Reactions were initiated by the addition of 10 µL of peptide substrate dissolved in DMSO. Aliquots (35 µL) of the reaction mixture were removed at 4 min time intervals up to 20 min and quenched into 1 mL of 3:2 chloroform:methanol + 200 µL of 4 mM $MgCl_2$. The aqueous layer was extracted and the organic layer was washed twice with 300 µL of pure solvent upper phase. The aqueous layers were combined, mixed with 5 mL of EcoLite scintillation fluid (MP Biomedicals), and subjected to scintillation counting. Eight reactions were set up in parallel in each assay, and all assays were carried out in duplicate or better.

Glycosylation of Protein Substrates Using PglB.

The reaction protocol is similar to the above procedure for peptide glycosylation, with the modification that 50 µL of PglB membrane fraction was used to glycosylate 13 µL of a 0.75 mM stock of Im7 protein acceptor in buffer P. The final concentrations for the protein and the glycan were 50 µM and 1.5 µM respectively. The reaction was left shaking at room temperature for 3 hours before being spun at high speed to pellet the membrane fraction. The supernatant was

applied to a Ni-NTA spin cartridge (Qiagen), washed and eluted according to manufacturer instructions. A 100 μL aliquot of the elution fraction was injected onto an analytical reverse-phase C_{18} HPLC column and eluted under a standard water/acetonitrile gradient. Fractions were collected every minute, mixed with EcoLite scintillation fluid (MP Biomedicals), and subjected to scintillation counting. Experiments using eukaryotic glycosyl donors were carried out using the same general procedure. For mass spectral analysis, a 5 nmol aliquot of unradiolabeled sugar donor was used as the glycosyl donor.

Circular Dichroism Evaluation of Acceptor Proteins.

Far-UV CD spectra were acquired on an Aviv Model 202 spectropolarimeter (Aviv Biomedical) using a 1 mm path length cell and a protein concentration of 250 μg/mL in buffer C [50 mM sodium phosphate, 400 mM sodium sulfate] at 25°C.

References

1. Burda P, Aebi M (1999) The dolichol pathway of N-linked glycosylation. *Biochim. Biophys. Acta.* **1426**: 239-57.

2. Hart GW, Brew K, Grant GA, Bradshaw RA, Lennarz WJ (1979) Primary structural requirements for the enzymatic formation of the N-glycosidic bond in glycoproteins. Studies with natural and synthetic peptides. *J. Biol. Chem.* **254**: 9747-53.

3. Bause E (1983) Structural requirements of N-glycosylation of proteins. Studies with proline peptides as conformational probes. *Biochem. J.* **209**: 331-6.

4. Heijne GV, Gavel Y (1990) Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering. *Protein. Eng.* **3**: 433-442.

5. Jones J, Krag SS, Betenbaugh MJ (2005) Controlling N-linked glycan site occupancy. *Biochim. Biophys. Acta.* **1726**: 121-37.

6. Young NM, Brisson JR, Kelly J, Watson DC, Tessier L, Lanthier PH, Jarrell HC, Cadotte N, St Michael F, Aberg E, Szymanski CM (2002) Structure of the N-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, *Campylobacter jejuni*. *J. Biol. Chem.* **277**: 42530-9.

7. Linton D, Dorrell N, Hitchen PG, Amber S, Karlyshev AV, Morris HR, Dell A, Valvano MA, Aebi M, Wren BW (2005) Functional analysis of the *Campylobacter jejuni* N-linked protein glycosylation pathway. *Mol. Microbiol.* **55**: 1695-703.

8. Wacker M, Linton D, Hitchen PG, Nita-Lazar M, Haslam SM, North SJ, Panico M, Morris HR, Dell A, Wren BW, Aebi M (2002) N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science* **298**: 1790-3.

9. Szymanski CM, Yao R, Ewing CP, Trust TJ, Guerry P (1999) Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. *Mol. Microbiol.* **32**: 1022-30.

10. Glover KJ, Weerapana E, Numao S, Imperiali B (2005) Chemoenzymatic synthesis of glycopeptides with PglB, a bacterial oligosaccharyl transferase from *Campylobacter jejuni*. *Chem. Biol.* **12**: 1311-5.

11. Weerapana E, Glover KJ, Chen MM, Imperiali B (2005) Investigating bacterial N-linked glycosylation: synthesis and glycosyl acceptor activity of the undecaprenyl pyrophosphate-linked bacillosamine. *J. Am. Chem. Soc.* **127**: 13766-7.

12. Nita-Lazar M, Wacker M, Schegg B, Amber S, Aebi M (2005) The N-X-S/T consensus sequence is required but not sufficient for bacterial N-linked protein glycosylation. *Glycobiology* **15**: 361-7.

13. Linton D, Allan E, Karlyshev AV, Cronshaw AD, Wren BW (2002) Identification of N-acetylgalactosamine-containing glycoproteins PEB3 and CgpA in *Campylobacter jejuni*. *Mol. Microbiol.* **43**: 497-508.

14. Bause E, Legler G (1981) The role of the hydroxy amino acid in the triplet sequence Asn-Xaa-Thr(Ser) for the N-glycosylation step during glycoprotein biosynthesis. *Biochem. J.* **195**: 639-44.

15. Kasturi L, Chen HG, ShakinEshleman SH (1997) Regulation of N-linked core glycosylation: Use of a site-directed mutagenesis approach to identify Asn-Xaa-Ser/Thr sequons that are poor oligosaccharide acceptors. *Biochem. J.* **323**: 415-419.

16. Imperiali B, Hendrickson TL (1995) Asparagine-linked glycosylation: specificity and function of oligosaccharyl transferase. *Bioorg. Med. Chem.* **3**: 1565-78.

17. Imperiali B, Shannon KL (1991) Differences between Asn-Xaa-Thr-containing peptides: a comparison of solution conformation and substrate behavior with oligosaccharyltransferase. *Biochemistry* **30**: 4374-80.

18. Imperiali B, Shannon KL, Rickert KW (1992) Role of Peptide Conformation in Asparagine-Linked Glycosylation. *J. Am. Chem. Soc.* **114**: 7942-7944.

19. Imperiali B, Spencer JR, Struthers MD (1994) Structural and Functional-Characterization of a Constrained Asx-Turn Motif. *J. Am. Chem. Soc.* **116**: 8424-8425.

20. Shakin-Eshleman SH, Spitalnik SL, Kasturi L (1996) The amino acid at the X position of an Asn-X-Ser sequon is an important determinant of N-linked core-glycosylation efficiency. *J. Biol. Chem.* **271**: 6363-6.

21. Weerapana E, Imperiali B (2006) Asparagine-linked protein glycosylation: From eukaryotic to prokaryotic systems. *Glycobiology* **16**: 91-101.

22. Pless DD, Lennarz WJ (1977) Enzymatic conversion of proteins to glycoproteins. *Proc. Natl. Acad. Sci. U.S.A.* **74**: 134-8.

23. Capaldi AP, Kleanthous C, Radford SE (2002) Im7 folding mechanism: misfolding on a path to the native state. *Nat. Struct. Biol.* **9**: 209-16.

24. Friel CT, Capaldi AP, Radford SE (2003) Structural analysis of the rate-limiting transition states in the folding of Im7 and Im9: similarities and differences in the folding of homologous proteins. *J. Mol. Biol.* **326**: 293-305.

25. Hackenberger CP, Friel CT, Radford SE, Imperiali B (2005) Semisynthesis of a glycosylated Im7 analogue for protein folding studies. *J. Am. Chem. Soc.* **127**: 12882-9.

26. Chak KF, Safo MK, Ku WY, Hsieh SY, Yuan HS (1996) The crystal structure of the immunity protein of colicin E7 suggests a possible colicin-interacting surface. *Proc. Natl. Acad. Sci. U.S.A.* **93**: 6437-6442.

CHAPTER 6

POLYISOPRENE SPECIFICITY IN THE PGL PATHWAY

Acknowledgements

The isolation of the polyisoprenols used in this study was carried out by Ewa Ciepichal from the Polish Academy of Science.

The PglC/PglA coupled assay of polyisoprenyl phosphates was carried out by Dr. Eranthie Weerapana from our laboratory.

The mass spectral characterization of polyisoprenyl diphosphate-disaccharides was carried out by Dr. Christopher Reid and Dr. Jacek Stupak from the Canadian National Research Council of Canada.

A significant portion of this chapter was published in:

Chen MM, Weerapana E, Ciepichal E, Stupak J, Reid CW, Swiezewska E, Imperiali B (2007) Polyisoprenol specificity in the *Campylobacter jejuni* N-linked glycosylation pathway. *Biochemistry* **46**: 14342-8.

In N-linked glycosylation, a glycan is assembled on a polyisoprenol-carrier by a series of glycosyltransferases. The completed glycan is then transferred by an oligosaccharyltransferase (OT) from the polyisoprenol-carrier to the acceptor protein (1). The substrates for both of the glycosyltransferases and the oligosaccharyltransferase are amphiphilic, with the polyisoprene being hydrophobic and embedded within the membrane, while the glycan is hydrophilic and is present in solution at the membrane interface (**Figure 6-1**). The recognition of these amphiphilic substrates by their membrane-associated enzymes is a poorly understood phenomenon due to the difficulties of studying these enzymes *in vitro* and of obtaining the appropriate polyisoprenol-linked substrates. One particular question of interest is whether the polyisoprenol-carrier plays a simple physical role as a hydrophobic membrane anchor to keep the glycan where the enzymes are localized, or whether they interact with the enzymes and play a more specific role in substrate recognition.

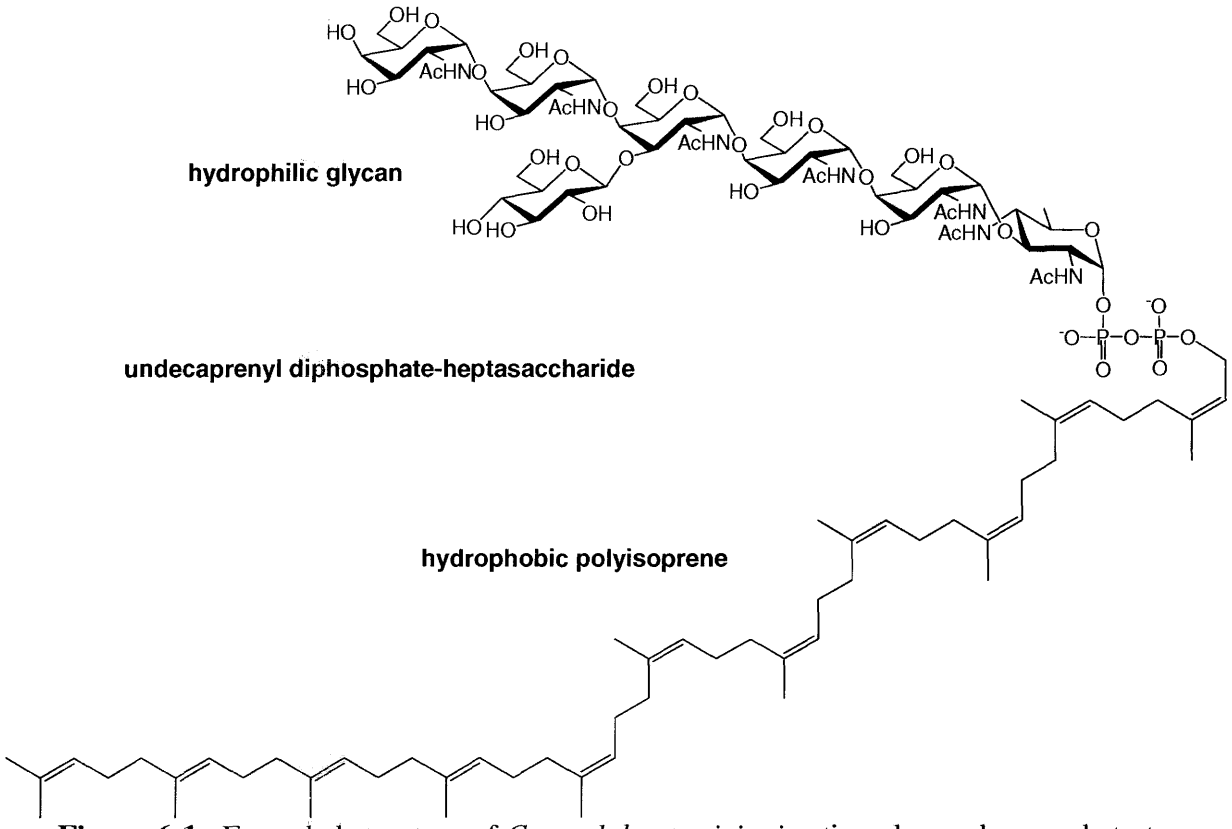


Figure 6-1. Extended structure of *Campylobacter jejuni* native glycan donor substrate.

The recently characterized N-linked protein glycosylation (Pgl) pathway from the gram-negative bacterium *Campylobacter jejuni* offers a unique opportunity to study the interaction between such enzymes and their polyisoprenol-linked substrates (**Figure 6-2**) (1,2). Firstly, the Pgl pathway offers a topologically diverse set of enzymes ranging from those that have numerous predicted transmembrane domains to those with none, and each of these enzymes can be overexpressed and purified for in vitro analysis. Secondly, as described in Chapter 4, we have the ability to synthesize highly-pure, specifically radiolabeled substrates and substrate analogs, which can be used to probe any enzyme within the Pgl pathway (3,4). In this study, we will carry out the first systematic study to investigate whether the structure of the polyisoprenol-carriers affects enzyme recognition, and if so, whether these effects are correlated with the degree of membrane association of the enzyme.

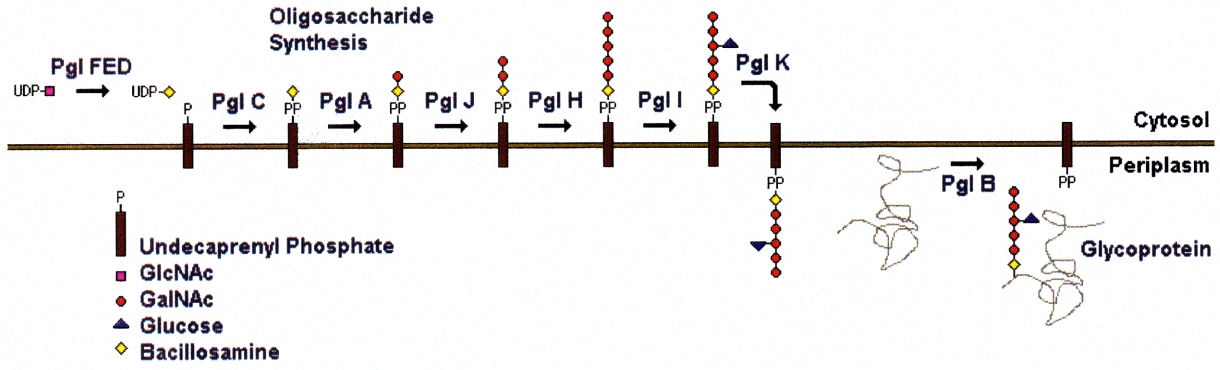


Figure 6-2. *C. jejuni* N-linked glycosylation pathway.

The three candidate enzymes from the Pgl pathway for this study include PglC, PglJ, and PglB, which differ significantly in the degree of membrane association and localization within the pathway (**Figure 6-4**). The glycosylphosphotransferase PglC transfers 2,4-diacetamido-bacillosamine phosphate (from here on referred to as simply bacillosamine phosphate) to

undecaprenyl phosphate to afford the first polyisoprenol-linked intermediate in the pathway (**Figure 6-3**) (4). PglC is predicted to contain a single N-terminal transmembrane domain of 22 amino acids (TMHMM, ExPASY) and a large cytoplasmic domain. An intermediate step in the pathway is catalyzed by PglJ, a glycosyltransferase, which transfers a GalNAc to undecaprenyl diphosphate-bacillosamine (3). PglJ is not predicted to include any transmembrane domains (TMHMM, ExPASY), but may contain an N-terminal hydrophobic domain, which is postulated to interact with the bacterial membrane.

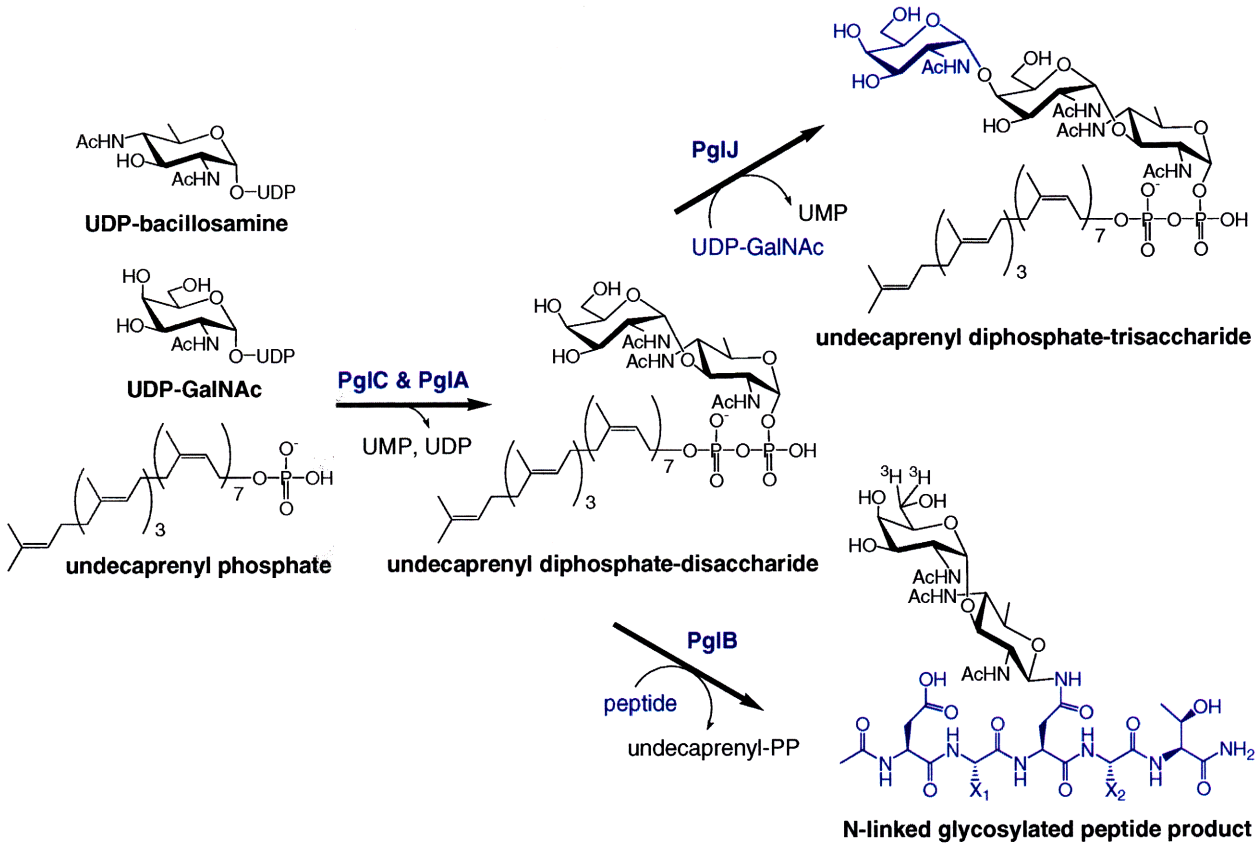


Figure 6-3. Summary of PglC, PglA, PglJ, and PglB enzyme reactions used in this study.

The final step in the pathway is catalyzed by the OT, PglB, which transfers the completed heptasaccharide from the undecaprenyl diphosphate-carrier to the protein acceptor (**Figure 6-3**)

(5). PglB comprises 10-12 predicted transmembrane domains from residues 1-425 (TMHMM, ExPASY) and a C-terminal periplasmic domain from residues 426-703. These three enzymes conveniently provide a panel of proteins that span the spectrum of membrane association, ranging from zero to multiple integral membrane regions (**Figure 6-4**). Since the polyisoprene is embedded within the membrane, one might hypothesize that the more transmembrane domains an enzyme contains, the more it might interact with the polyisoprene, and the more discriminating it will be towards the polyisoprene structure.

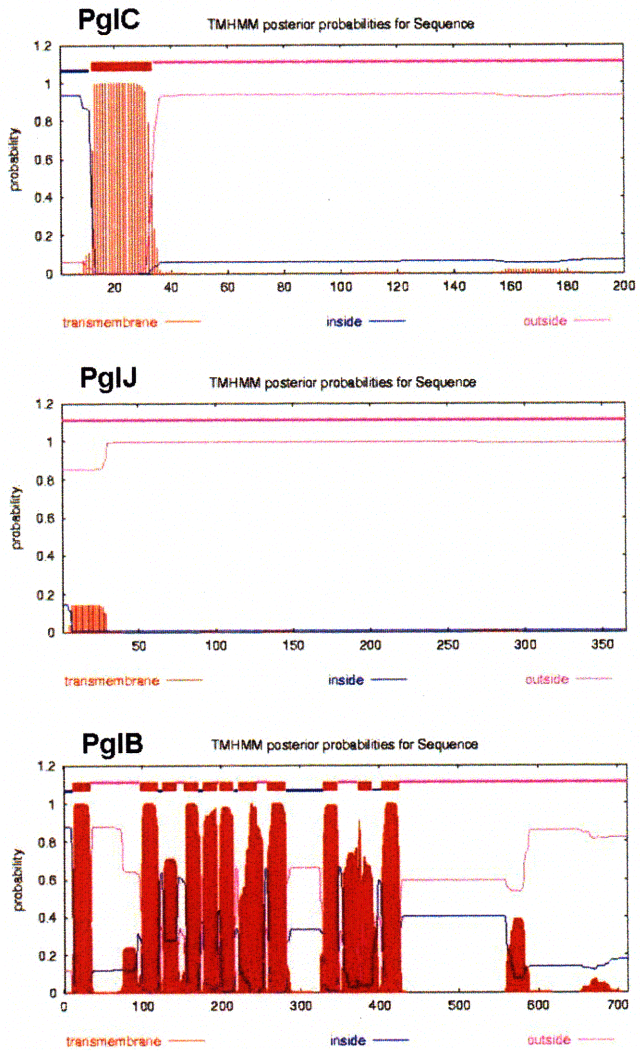


Figure 6-4. TMHMM, ExPASY predicted transmembrane domains of PglC, PglJ, and PglB.

6-1. Synthesis of polyisoprenyl-linked substrates

In order to explore the determinants that mediate the enzyme interaction with their amphiphilic substrate, a set of polyisoprenols of varying length, degrees of saturation, and double-bond geometry were obtained from our collaborators at the Polish Academy of Science who derived the pure polyisoprenols from natural plant sources (**Figure 6-5**) (6,7). The native polyisoprenol in the *C. jejuni* N-linked glycosylation pathway is undecaprenol, which contains 11 unsaturated isoprene units, with *cis* (*Z*) double-bond geometry at the α-isoprene unit and an array of *trans* (*E*) and *cis* isoprene units in the remainder of the structure. Although they share the same name, our undecaprenol was isolated from the plant *Magnolia kobus* which contains a 7:3 *cis/trans* ratio, while the undecaprenol found in bacterial sources is slightly different with an 8:2 ratio. To determine the effect of polyisoprenol length on enzyme activity, we obtained prenol-9 and prenol-8, which are very similar to undecaprenol in saturation and double bond geometry, but contain two and three fewer isoprene units respectively. To determine the effect of double-bond geometry, we obtained solanesol and geranylgeraniol, which consist exclusively of *trans* double-bonds, rather than a mixture of *cis* and *trans* isomers. Prenol-9 provides a convenient length control for solanesol because these two polyisoprenols contain the same total number of isoprene units. Finally, to determine the effect of the α-isoprene saturation, we obtained dihydroprenol-11 and the dolichols, which are respectively either the same length as undecaprenol or longer, and are both saturated at the α-isoprene unit. This results in a sp^3-hybridized tetrahedral carbon adjacent to the alcohol rather than a planar sp^2-hybridized carbon center of the native undecaprenol. The dolichols are particularly interesting because they constitute part of the native substrates in the eukaryotic N-linked glycosylation pathways (8).

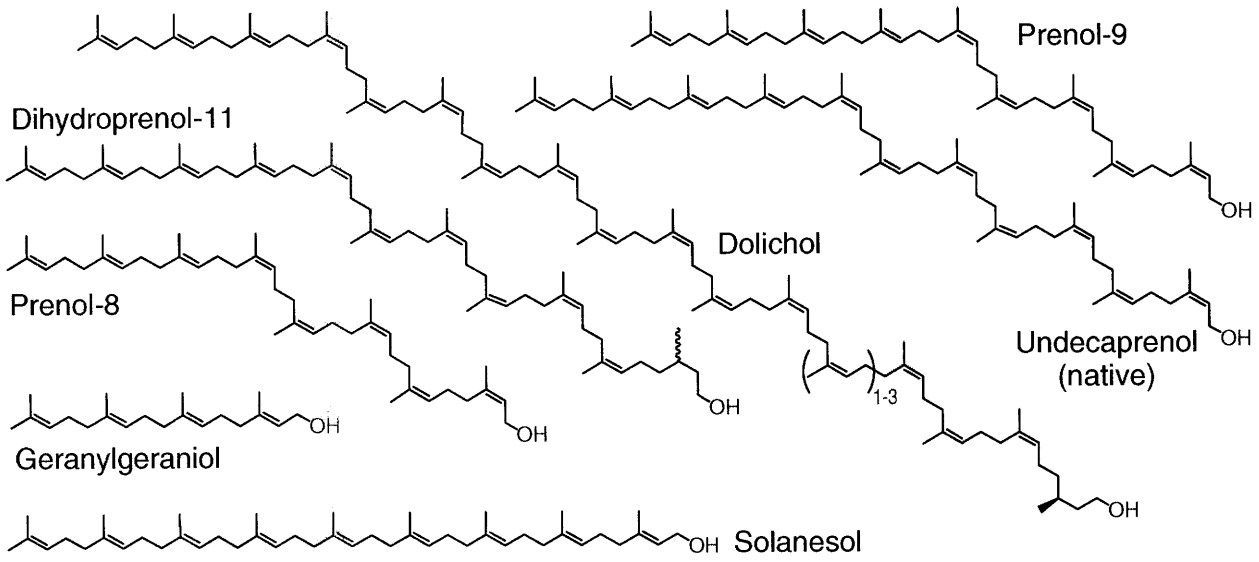


Figure 6-5. Polyisoprenols used in this study.

The polyisoprenols were incorporated into substrates for the enzymes PglC, PglJ and PglB using a chemoenzymatic approach (**Figure 6-6**). To make the appropriate substrates for PglC, each polyisoprenol was chemically phosphorylated (9,10) to afford the polyisoprenyl phosphate. To makes the appropriate substrates for PglB and PglJ, these polyisoprenyl-phosphates were converted to the respective polyisoprenyl diphosphate-bacillosamine-[^{3}H]GalNAc using PglC and the glycosyltransferase PglA as described in Chapters 4 and 5. Although the native substrate for PglB is an undecaprenyl diphosphate-linked heptasaccharide donor, the bacillosamine-[^{3}H]GalNAc disaccharide was used in our in vitro assays of PglB because of its easier synthesis and better solubility properties as described in Chapter 5 (5). The overall yield of the disaccharide synthesis using PglC and PglA depended on the identity of the polyisoprenyl phosphates. By using an excess UDP-bacillosamine and UDP-GalNAc to drive the reaction, typical yields ranged between 20-70% for a 50 nmol scale preparation. The disaccharide products were separated from the unreacted polyisoprenyl phosphates using normal phase HPLC

and characterized by ESI-MS. The [^{3}H] radiolabel offers a convenient means to accurately quantify both substrates and products in the low micromolar concentration range. This chemoenzymatic synthesis once again demonstrates the versatility of our approach for accessing polyisoprenol-linked glycan substrates for biochemical studies.

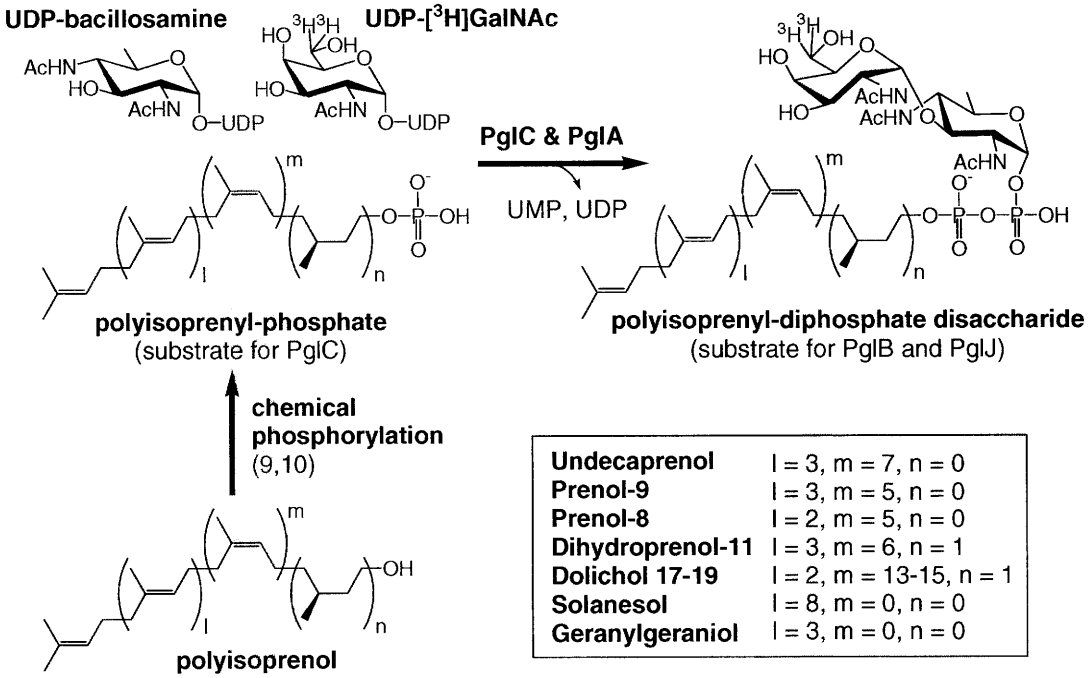


Figure 6-6. Synthesis of polyisoprenyl phosphate and polyisoprenyl diphosphate-disaccharide as substrates for PglC, PglB, and PglJ.

6-2. Comparison of polyisoprenyl-substrate analogues with PglB

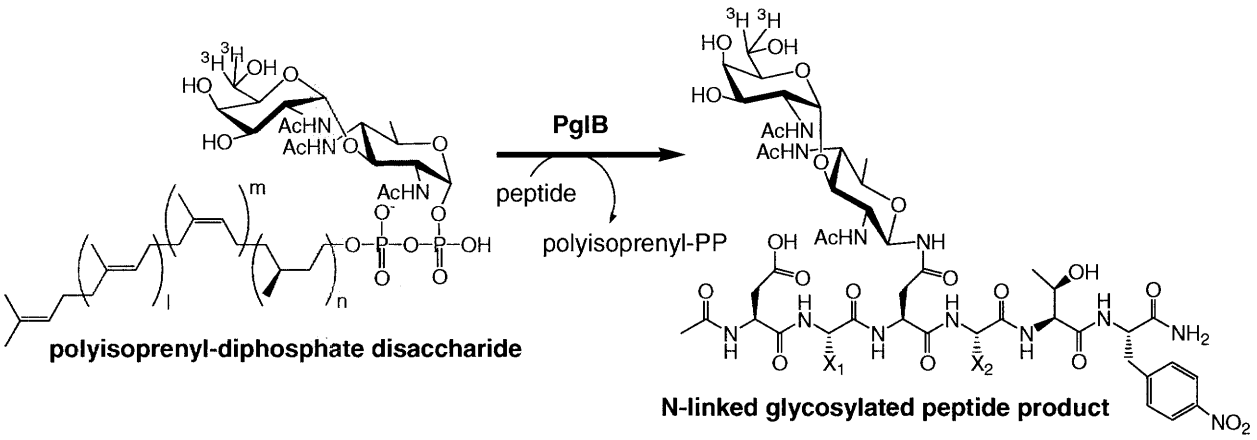


Figure 6-7. PglB in vitro assay reaction.

With purified substrates and detergent solubilized enzymes, steady-state kinetic assays were performed to determine the specificity of PglC, PglJ and PglB for each of the different polyisoprenol-containing substrate analogs. To assay for PglB activity, the transfer of radiolabeled glycan from the organic soluble polyisoprenyl diphosphate donor to the aqueous soluble peptide acceptor was monitored (**Figure 6-7**). The peptide acceptor used in this study was the hexapeptide Ac-DFNVT-(*p*NF)-NH_2 where *p*NF is *para*-nitrophenylalanine described in Chapter 5 as providing a convenient chromophore for peptide concentration quantification (11). The peptide was used at saturating concentrations and the amount of the polyisoprenyl diphosphate-linked disaccharide substrates was varied to afford the results shown. The reaction rates with the prenol-9 and prenol-8 analogs are very similar to that of the native undecaprenol (**Figure 6-8B**). This suggests that PglB is not affected by the presence of a slightly truncated form of undecaprenol. The use of dihydroprenol-11 and dolichols, both of which include a saturated terminal α-isoprene unit, resulted in a significant decrease in PglB activity suggesting that PglB is sensitive to the degree of saturation proximal to the site of enzymatic action.

Finally, the solanesol-linked substrate demonstrated the lowest activity with PglB, suggesting that the all-*trans* geometry also disrupts interactions between PglB and the polyisoprenol moiety. Both the K_M and V_{max} are affected in the case of these non-natural substrates (**Table 6-1**).

Table 6-1. Kinetic parameters of polyisoprenyl-substrate with Pgl enzymes. PglC and PglJ initial rates refer to initial reaction rates at 60 μM polyisoprenyl phosphate and 0.5 μM polyisoprenyl diphosphate-disaccharide concentrations respectively. PglB K_m values refer to that of the polyisoprenyl-substrate at saturating peptide acceptor concentration. n.d. = not determined.

Polyisoprenol	PglC	PglJ	PglB		
	Initial Rate nM/min	Initial Rate nM/min	V_{max} nM/min	K_m μM	V_{max}/K_m $min^{-1} \times 10^3$
Prenol-11 (undecaprenol)	210.6 ± 18.1	1.83 ± 0.20	63.2 ± 4.6	2.63 ± 0.36	24.0 ± 5.0
Prenol-9	214.4 ± 24.4	1.66 ± 0.11	56.0 ± 5.0	2.97 ± 0.50	18.9 ± 4.8
Prenol-8	183.8 ± 6.2	n.d	64.4 ± 6.9	2.77 ± 0.56	23.2 ± 7.2
Dihydroprenol-11	57.5 ± 5.6	0.67 ± 0.05	24.4 ± 2.0	5.87 ± 0.82	4.2 ± 0.9
Dihydroprenol-17-19 (dolichols)	50.6 ± 5.1	0.54 ± 0.03	16.6 ± 1.3	4.23 ± 0.53	3.9 ± 0.8
all-trans Prenol-9 (solanesol)	3.8 ± 2.0	0.54 ± 0.07	8.5 ± 0.8	4.06 ± 0.68	2.1 ± 0.6
all-trans Prenol-4 (geranylgeraniol)	0.6 ± 1.0	n.d	n.d.	n.d.	n.d.

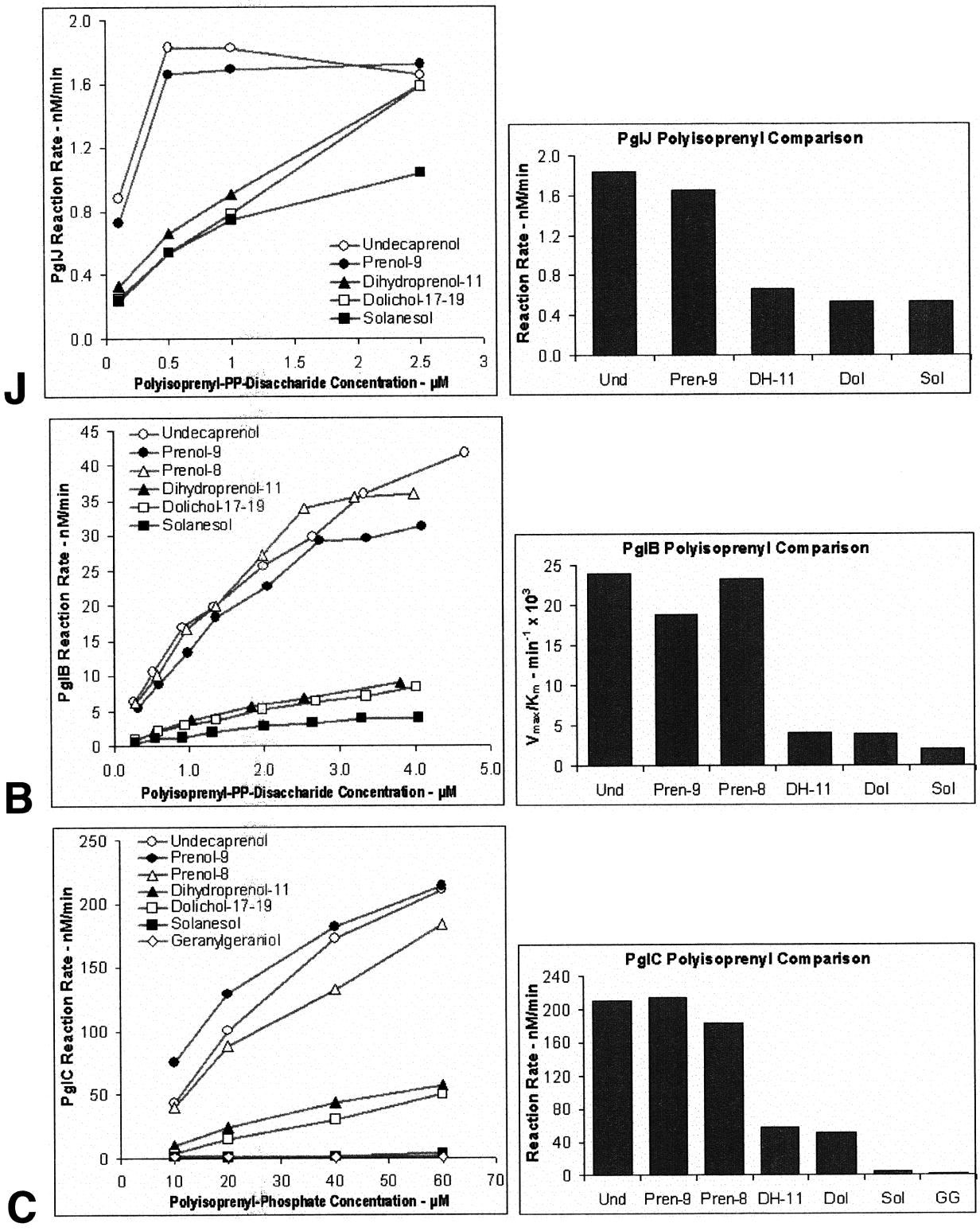


Figure 6-8. Comparison of enzyme reaction rates at varying concentrations of polyisoprenyl-linked substrates: **J**) glycosyltransferase PglJ, **B**) oligosaccharyltransferase PglB, and **C**) glycosylphosphotransferases PglC. Adjacent bar graph results are described in **Table 6-1**. All assays were carried out in duplicate or better with representative data set shown.

6-3. Comparison of polyisoprenyl-substrate analogues with PglC

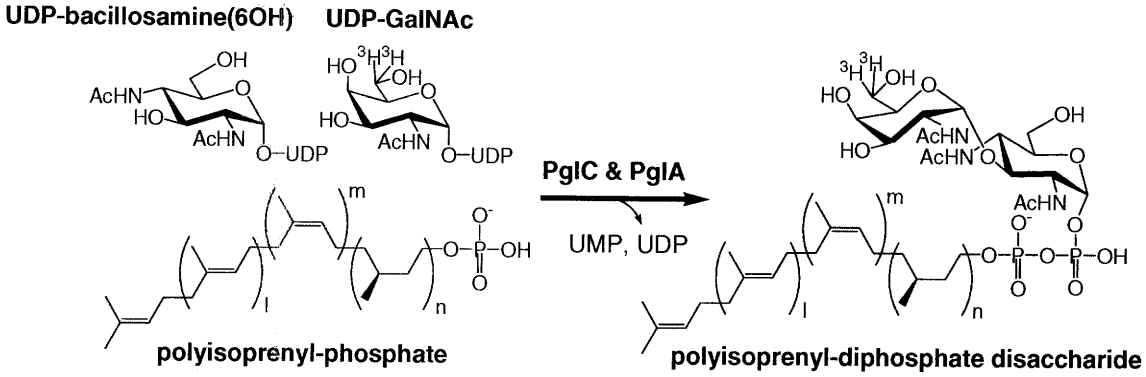


Figure 6-9. PglC in vitro assay reaction.

In the same manner, PglC was assayed against each of the polyisoprenyl phosphates by Dr. Eranthie Weerapana from our laboratory (4). The UDP-bacillosamine(6OH) analog was used as the glycosyl-phosphate donor since it is known to be accepted by PglC with comparable efficiency as the native UDP-bacillosamine (4) and was more readily available at that time because of its shorter chemical synthesis. To assay for PglC activity, a coupled assay involving PglA (**Figure 6-9**) (4) was used because of the lack of a radiolabeled UDP-bacillosamine to directly assay PglC at the time the assays were carried out. A large excess of PglA was used to ensure that reaction rates reflected the rate limiting PglC step, and that the rates measured did not change when additional PglA was added. The assay for PglC activity monitored the transfer of radioactivity from the aqueous soluble UDP-[^{3}H]GalNAc to the organic soluble polyisoprenyl diphosphate-substrate. **Figure 6-8C** illustrates the reaction rates using each of the different polyisoprenyl phosphates as substrates for PglC. Similar to the substrate selectivity demonstrated by PglB, PglC accepts prenol-8 and prenol-9 with equal efficiencies to the native undecaprenol-based substrate. Dolichols-, and dihydroprenol-11, substrates were accepted with

approximately four-fold decrease in activity, suggesting that PglC is sensitive to the saturation of the α-isoprene unit. Finally, the all-*trans* solanesyl and geranylgeranyl phosphates were extremely poor substrates for PglC suggesting that the isoprene-geometry plays a critical role in substrate recognition.

6-4. Comparison of polyisoprenyl-substrate analogues with PglJ

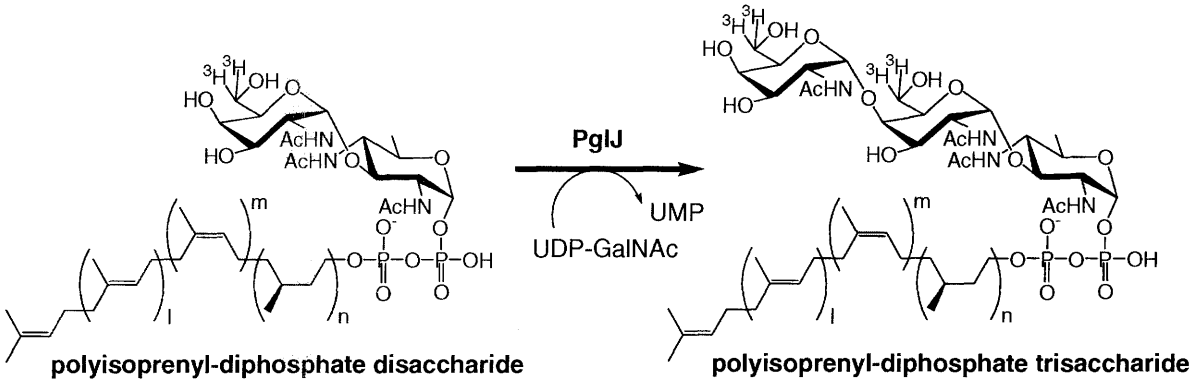


Figure 6-10. PglJ in vitro assay reaction.

PglJ is the peripheral membrane protein that transfers the third sugar in the heptasaccharide biosynthesis. The PglJ enzyme samples used in this study were purified to homogeneity using Ni-NTA affinity chromatography as previously described (3). The same purified radiolabeled disaccharide substrates used in the PglB assay above, were also used to assay PglJ activity. In this case, since the disaccharide acceptor is already radiolabeled, we simply measured the increase in radioactivity resulting from the transfer of an additional GalNAc residue from UDP-[^{3}H]GalNAc (**Figure 6-10**). Similar to PglB and PglC, PglJ accepts the prenol-9 substrate with equal efficiencies to the native undecaprenol-linked substrate (**Figure 6-8J**). The non-natural

dihydroprenol-11, dolichols and solanesol-linked substrates are also accepted by PglJ but with lower reaction rates compared to the corresponding undecaprenol and prenol-9 substrates.

Our studies reveal that each of the three enzymes from the Pgl pathway exhibit distinct preferences for the structural features of the polyisoprenols within their target substrates. In comparison to the native undecaprenol-based substrates, polyisoprenol-substrates of shorter length, such as prenol-9 and prenol-8, displayed similar turnover, indicating that all three enzymes do not have stringent specificity for the number of isoprene units; the C40 prenol-8 is virtually identical in activity to the C55 native undecaprenol. Although solanesol is of the same length as prenol-9, its incorporation into substrates for each of the enzymes proved highly deleterious, most likely due to the unfavorable all-*trans* double bond geometry. Furthermore, dihydroprenols such as dolichols, being saturated at the α-isoprene position, were also poorly accepted in comparison to the unsaturated counterparts. This clear specificity with respect to the polyisoprenol-carrier, especially towards the features at α-isoprene units, is consistent with the hypothesis that membrane-associated hydrophobic proteins interact extensively with their lipophilic substrates. Intriguingly, there was no clear correlation between the stringency of the substrate specificity and the number of predicted transmembrane domains possessed by each enzyme.

We considered the possibility that the differences observed in V_{max} and K_M do not actually reflect differences in catalytic efficiencies or their binding affinities to the enzymes, but rather a function of substrate solubility as well as, in the case of PglB and PglC, enzyme accessibility which is limited by the permeability of the cell membrane. However, the concentration of

detergent used in the assay conditions was sufficient for fully dissolving the cell membrane, and therefore the interaction of the polyisoprenyl-linked substrates should not be influenced or limited by the membrane structure (12). Furthermore, the detergent is known to form mixed micelles with the substrate, such that each polyisoprenyl-linked substrate should be effectively solubilized (13).

These studies concur with a previous report on the N-linked glycosylation pathway of *Saccharomyces cerevisiae*, whereby polyisoprenyl-specificity was found to be maintained by both Alg7 and the OT, the two eukaryotic enzymes that catalyze the analogous reactions to PglC and PglB (14). However, since Alg7 did not accept any α-unsaturated polyisoprenyl phosphates or the all-*trans* solanesyl phosphate, the corresponding OT substrates could not be accessed and therefore the study could not be pursued to the level of detail presented herein for the *C. jejuni* enzymes. To our knowledge, the specificity of OT towards α-unsaturated and all-*trans* polyisoprenols has not been previously tested in any other systems. A simplified generalization has been that α-unsaturated polyisoprenols are primarily found in bacteria and plants, while dihydropolyisoprenols are present in mammalian and yeast cells (6). Therefore, many examples of glycosyltransferases from eukaryotic and prokaryotic sources have been found to reflect this specificity towards their native type of polyisoprenol-carrier (15-17). In this case, the in vitro radioactivity-based assay allowed us to takes advantage of the versatility of a chemoenzymatic synthesis to evaluate unnatural substrates for the bacterial OT. Interestingly, all three enzymes appear capable of accepting the eukaryotic dolichol-based carrier in vitro albeit with lower reaction rate.

Conclusion

Using the N-linked glycosylation pathway of *C. jejuni* as a model system for the study of how membrane-associated enzymes interact with their lipophilic substrates, we selected three enzymes from this pathway, PglC, PglJ, and PglB, which have varying degrees of membrane-association. A set of polyisoprenols varying in length, degrees of saturation, and double-bond geometry were obtained from natural sources and converted into the respective substrates for these three enzymes using a chemo-enzymatic synthesis. We found that each enzyme displayed the same specificity and preferences towards the set of polyisoprenyl-linked substrates, whereby the overall length of the polyisoprene did not affect enzyme activity, but changes to the structure of the α-isoprene led it being a poorer substrate for the enzyme. This specificity towards the polyisoprenol portion of the substrate by all three enzymes suggests that the polyisoprenol does not play merely a physical role as a hydrophobic membrane anchor, but interacts with the enzymes from the pathway and is involved in substrate recognition.

Experimental

UDP-sugar Donors.

UDP-bacillosamine was synthesized enzymatically from UDP-GlcNAc using the PglF, PglE, and PglD enzymes from the *C. jejuni* pathway as described previously in Chapter 4 (18). The preparation of UDP-bacillosamine(6OH), which is a synthetic intermediate in the chemical synthesis of UDP-bacillosamine was carried out as reported previously (4,19). Radiolabeled UDP-[^{3}H]GalNAc was purchased from American Radiolabeled Chemicals Inc.

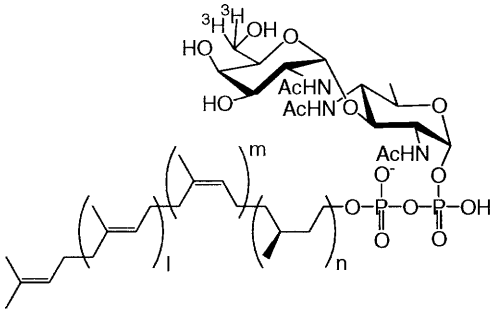
Polyisoprenol Isolation & Phosphorylation.

Solanesol and geranylgeraniol were purchased from Sigma-Aldrich. All remaining polyisoprenols were isolated from diverse plant sources (7) and purified from the natural extracts by chromatographic methods (20). Undecaprenol was from *Magnolia kobus* leaves, prenol-9 was from the *Tilia cordata* leaves, prenol-8 was from the wood of *Betula veruccosa,* and prenols-17-19 were from *Ginkgo biloba* leaves. Hydrogenation of undecaprenol resulted in a racemic mixture of dihydroprenol-11 (21). Dolichols-17-19 were synthesized via asymmetric hydrogenation of prenols-17-19 (22). Undecaprenol, dolichols, solanesol, and geranylgeraniol were chemically phosphorylated using phosphoramidite chemistry (9,10). Prenol-9, prenol-8, were phosphorylated with phosphoric acid and trichloroacetonitrile (23). Dihydroprenol-11 was phosphorylated with phosphorus oxychloride (24). The purity of the polyisoprenyl-phosphates was evaluated by thin layer chromatography and determined to be >95% in all cases.

Enzyme Purification.

The expression and purification of PglA (3), PglJ (3), PglC (4), and PglB (5) enzymes have been previously described in detail. For this study, both PglA and PglJ were further subjected to an additional chromatographic purification on a Sephacryl 300 gel filtration column (GE Healthcare). PglA and PglJ were evaluated to be 95% pure by SDS PAGE, while PglC and PglB were used as a semi-pure membrane fraction necessary for preserving enzyme stability.

Enzymatic synthesis of Radiolabeled Polyisoprenyl Diphosphate Disaccharide Donors.



To an eppendorf tube containing 50 nmol of the polyisoprenyl-phosphate was added 6 µL DMSO and 14 µL of 14.3% (v/v) Triton X-100. After vortexing and sonication (water bath), 42.5 µL of H_2O, 5.5 µL of 1 M Tris-acetate (pH 8.0), 2 µL of 1 M $MgCl_2$, 20 µL of 5 mM UDP-Bac2,4diNAc, 20 µL of 2.5 µM UDP-GalNAc (specific activity = 152 nCi $nmol^{-1}$), 70 µL of PglA, and 40 µL of PglC were added. The concentrations of enzyme stock solutions used were approximately 500 mg mL^{-1} and 250 mg mL^{-1} for PglA (3) and PglC (4) respectively. Reactions were left shaking at room temperature for 3 hours then quenched by the addition of 800 µL 2:1 chloroform:methanol and 100 µL of pure solvent upper phase (PSUP; 3% chloroform, 49% methanol, and 48% water with 100 mM KCl). After brief vortexing, the layers were allowed to separate and the aqueous layer was removed. The organic layer was washed three times with 200 µL of PSUP, dried under a stream of nitrogen, and then redissolved in 100 µL of 4:1

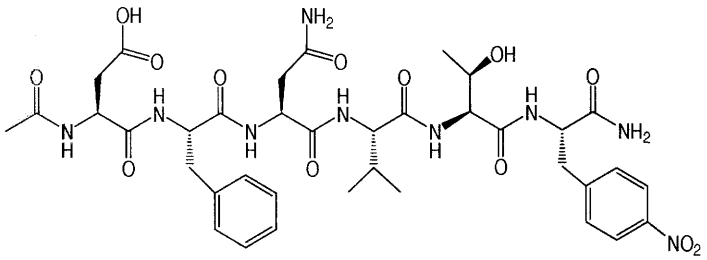
chloroform:methanol. The crude product was then purified by HPLC on a Microsorb 60 Si normal phase column (Varian), eluting with a linear gradient of 100% solvent C (4:1 chloroform/methanol) to 100% solvent D (10:10:3 chloform/methanol/2 M ammonium acetate pH 7.2) over 50 min at a flow rate of 1 mL/min. Fractions containing significant amount of radiolabeled product were combined and immediately aliquoted, dried, and stored at -80 °C until needed.

PglJ Radioactivity Assay.

To a tube containing a specified amount of dried radiolabeled polyisoprenyl diphosphate Bac2,4diNAc-[^{3}H]GalNAc were added 10 µL DMSO and 7 µL of 1.43% (v/v) Triton X-100. After vortexing and sonication (water bath), 5 µL of 1 M Tris-acetate (pH 8.0), 1 µL of 1 M $MgCl_2$, 5µL of 50 mM DTT, 62 µL of H_2O, and 5 µL of 500 mg mL^{-1} PglJ were added. For negative control reactions (not shown), the PglJ samples were replaced by 50 mM Tris-acetate buffer (pH 8.0). Four reactions containing different amounts of polyisoprenyl diphosphate disaccharides were carried out in parallel. The reactions were initiated by the addition of 5 µL of 1 µM radiolabeled UDP-GalNAc (specific activity = 19.8 µCi $nmol^{-1}$, K_M(UDP-GalNAc) ~1 µM). All assays were carried out at 22 ± 1 °C, and 17 µL aliquots were taken at 2, 4, 6, and 8 min. Reactions were quenched by addition to an eppendorf tube containing 800 µL 2:1 chloroform/methanol and 200 µL of PSUP (25). After brief vortexing, the layers were allowed to separate and the aqueous layer was removed. The organic layer was washed two times with 100 µL of PSUP and dried under a stream of nitrogen. The residue was then redissolved in 200 µL of DMSO by vigorous vortexing followed by the addition of 5 mL of EcoLite scintillation fluid

(MP Biomedicals) and subjected to scintillation counting (5 min per sample). All assays were carried out in duplicate or better.

Peptide Synthesis.



The Ac-DFNVT-(*p*NF)-NH_2 acceptor peptides was synthesized by automated peptide synthesis on an ABI 431A peptide synthesizer (Applied Biosystems) using standard Fmoc-based peptide synthesis conditions on PAL-PEG-PS resin. This peptide was acetylated at the N-terminus, and included a *para*-nitrophenylalanine (*p*NF) at the C-terminus. The peptides were cleaved from the resin using a trifluoroacetic acid cocktail containing 2.5% water and 2.5% triisopropylsilane, purified to ≥95% purity by preparative C_{18} reverse-phase HPLC eluting with a linear gradient of 100% solvent A (water with 0.1% TFA) to 100% solvent B (acetonitrile with 0.1% TFA) over 30 min at a flow rate of 15 mL/min, and quantified using the UV absorbance of the *p*NF amino acid at 280 nm (ε = 12,500 $M^{-1}cm^{-1}$). The K_M of this peptide with PglB is 1.21 μM (11).

PglC and PglA Coupled Radioactivity Assay.

PglC was assayed based on procedures described elsewhere (4). To a tube containing a specified amount of dried polyisoprenyl-phosphate were added 3 μL DMSO and 7 μL of 14.3% (v/v) Triton X-100. After vortexing and sonication (water bath), 73 μL of H_2O, 4 μL of 1 M Tris-acetate pH 8.0, 1 μL of 1 M $MgCl_2$, 5 μL of PglA, and 2 μL of PglC. For negative control reactions (not shown), the PglA and PglC samples were replaced by 50 mM Tris-acetate buffer

pH 8.0. The concentrations of enzyme stocks used were 500 mg mL^{-1} and 250 mg mL^{-1} for PglA (3) and PglC (4) respectively. Four reactions containing different amounts of polyisoprenyl phosphate were carried out in parallel. The reactions were initiated by the addition of 5 μL of sugar mix consisting of 100 μM UDP-6-hydroxylBac2,4diNAc and 100 μM UDP-GalNAc (specific activity = 36.1 nCi $nmol^{-1}$, $K_{M(UDP\text{-}GalNAc)}$ ~2 μM, $K_{M(UDP\text{-}6OHBac)}$ ~4 μM) Reactions were carried out at 22 ± 1 °C, and 20 μL aliquots were taken at 2, 4, 6, and 8 min. Reactions were quenched by addition to a tube containing 400 μL 2:1 chloroform:methanol and 200 μL of pure solvent upper phase (PSUP; 3% chloroform, 49% methanol, and 48% water with 100 mM KCl). After being vortexed briefly the layers were allowed to separate and the aqueous layer was removed. The organic layer was washed two times with 400 μL of PSUP and dried under a stream of nitrogen. The residue was redissolved in 200 μL of Solvable (Perkin-Elmer) by vigorous vortexing followed by the addition of 5 mL of Formula 989 scitillation fluid (Perkin-Elmer). The tubes were allowed to rest for 1 hour and counted in a scintillation counter (5 min per sample). Eight reactions were set up in parallel in each assay, and all assays were carried out in duplicate.

PglB Radioactivity Assay.

PglB was assayed based on procedures described elsewhere (11). Briefly, to a tube containing a specified amount of dried radiolabeled [^{3}H]GalNAc-Bac2,4diNAc-PP-polyisoprenol (specific activity = 160 nCi $nmol^{-1}$) was added 10 μL of DMSO. Following vigorous vortexing and sonication (water bath) to resuspend the isoprene-based substrate, 100 μL of 2X assay buffer [280 mM sucrose, 2.4% Triton X-100 (v/v), 280 mM Hepes pH 7.5], 2 μL of 1 M $MnCl_2$, and 6 μL of PglB (5) membrane fraction containing approximately 50 ng of enzyme were added, and

the volume increased to 190 μL with water. Reactions were initiated by the addition of 10 μL of 2.0 mM peptide acceptor ($K_{M(peptide)}$ = 1.21 μM) dissolved in DMSO. The reactions were carried out at 22 ± 1 °C, and 35 μL aliquots of the reaction mixture were removed at 4 min time intervals up to 20 min and quenched into 1 mL of 3:2 chloroform/methanol and 200 μL of 4 mM $MgCl_2$. The aqueous layer was extracted and the organic layer was washed twice with 300 μL of PSUP. The aqueous layers were combined, mixed with 5 mL of EcoLite scintillation fluid (MP Biomedicals), and subjected to scintillation counting. Eight reactions were set up in parallel in each assay, and all assays were carried out in duplicate or better. Glycopeptide products have been previously confirmed for this assay (5).

Mass Spectral Characterization of Unnatural Polyisoprenyl Diphosphate Disaccharide Substrates.

Mass spectra were acquired using a 4000 QTrap mass spectrometer (Applied Biosystems / Sciex, Concord, ON, Canada). Capillary electrophoresis was performed using a Prince CE system (Prince Technologies, The Netherlands). The separation was obtained on a 90 cm length bare fused-silica capillary (365 μm OD x 50 μm ID) with CE-MS coupling using a liquid sheath-flow interface and isopropanol:methanol (2:1) as the sheath liquid. Precursor-ion scanning (m/z 79, PO_4) for the dolichol sample was done in the negative-ion mode, with a collision energy of 70 V. Samples were dissolved in 3:1 chloroform/methanol and an aqueous CE buffer consisting of 10 mM ammonium acetate was used for all experiments.

References

1. Weerapana E, Imperiali B (2006) Asparagine-linked protein glycosylation: From eukaryotic to prokaryotic systems. *Glycobiology* **16**: 91-101.

2. Linton D, Dorrell N, Hitchen PG, Amber S, Karlyshev AV, Morris HR, Dell A, Valvano MA, Aebi M, Wren BW (2005) Functional analysis of the *Campylobacter jejuni* N-linked protein glycosylation pathway. *Mol. Microbiol.* **55**: 1695-703.

3. Glover KJ, Weerapana E, Imperiali B (2005) *In vitro* assembly of the undecaprenylpyrophosphate-linked heptasaccharide for prokaryotic N-linked glycosylation. *Proc. Natl. Acad. Sci. U.S.A.* **102**: 14255-9.

4. Glover KJ, Weerapana E, Chen MM, Imperiali B (2006) Direct biochemical evidence for the utilization of UDP-bacillosamine by PglC, an essential glycosyl-1-phosphate transferase in the *Campylobacter jejuni* N-linked glycosylation pathway. *Biochemistry* **45**: 5343-50.

5. Glover KJ, Weerapana E, Numao S, Imperiali B (2005) Chemoenzymatic synthesis of glycopeptides with PglB, a bacterial oligosaccharyl transferase from *Campylobacter jejuni*. *Chem. Biol.* **12**: 1311-5.

6. Swiezewska E, Danikiewicz W (2005) Polyisoprenoids: structure, biosynthesis and function. *Prog. Lipid Res.* **44**: 235-58.

7. Wellburn AR, Hemming FW (1966) The occurrence and seasonal distribution of higher isoprenoid alcohols in the plant kingdom. *Phytochemistry* **5**: 969-975.

8. Lennarz WJ (1975) Lipid linked sugars in glycoprotein synthesis. *Science* **188**: 986-91.

9. Ye XY, Lo MC, Brunner L, Walker D, Kahne D, Walker S (2001) Better substrates for bacterial transglycosylases. *J. Am. Chem. Soc.* **123**: 3155-6.

10. Branch CL, Burton G, Moss SF (1999) An expedient synthesis of allylic polyprenyl phosphates. *Synth. Commun.* **29**: 2639-2644.

11. Chen MM, Glover KJ, Imperiali B (2007) From Peptide to Protein: Comparative Analysis of the Substrate Specificity of N-Linked Glycosylation in C. jejuni. *Biochemistry* **46**: 5579-85.

12. Bergman A, Mankowski T, Chojnacki T, De Luca LM, Peterson E, Dallner G (1978) Glycosyl transfer from nucleotide sugars to C85- and C55-polyprenyl and retinyl phosphates by microsomal subfractions and Golgi membranes of rat liver. *Biochem J* **172**: 123-7.

13. Low P, Peterson E, Mizuno M, Takigawa T, Chojnacki T, Dallner G (1986) Effectivity of dolichyl phosphates with different chain lengths as acceptors of nucleotide activated sugars. *Biosci Rep* **6**: 677-83.

14. Palamarczyk G, Lehle L, Mankowski T, Chojnacki T, Tanner W (1980) Specificity of solubilized yeast glycosyl transferases for polyprenyl derivatives. *Eur J Biochem* **105**: 517-23.

15. Rush JS, Shelling JG, Zingg NS, Ray PH, Waechter CJ (1993) Mannosylphosphoryldolichol-mediated reactions in oligosaccharide-P-P-dolichol biosynthesis. Recognition of the saturated alpha-isoprene unit of the mannosyl donor by pig brain mannosyltransferases. *J Biol Chem* **268**: 13110-7.

16. Mankowski T, Sasak W, Janczura E, Chojnacki T (1977) Specificity of polyprenyl phosphates in the in vitro formation of lipid-linked sugars. *Arch Biochem Biophys* **181**: 393-401.

17. Mankowski T, Sasak W, Chojnacki T (1975) Hydrogenated polyprenol phosphates - exogenous lipid acceptors of glucose from UDP glucose in rat liver microsomes. *Biochem Biophys Res Commun* **65**: 1292-7.

18. Olivier NB, Chen MM, Behr JR, Imperiali B (2006) *In Vitro* Biosynthesis of UDP-N,N'-Diacetylbacillosamine by Enzymes of the *Campylobacter jejuni* General Protein Glycosylation System. *Biochemistry* **45**: 13659-69.

19. Weerapana E, Glover KJ, Chen MM, Imperiali B (2005) Investigating bacterial N-linked glycosylation: synthesis and glycosyl acceptor activity of the undecaprenyl pyrophosphate-linked bacillosamine. *J. Am. Chem. Soc.* **127**: 13766-7.

20. Chojnacki T, Jankowski W, Mankowski T, Sasak W (1975) Preparative separation of naturally occurring mixtures of polyprenols on hydroxyalkoxypropyl-Sephadex. *Anal Biochem* **69**: 114-9.

21. Mankowski T, Jankowski W, Chojnacki T, Franke P (1976) C55-Dolichol: occurrence in pig liver and preparation by hydrogenation of plant undecaprenol. *Biochemistry* **15**: 2125-30.

22. Imperiali B, Zimmerman JW (1988) Synthesis of Dolichols Via Asymmetric Hydrogenation of Plant Polyprenols. *Tetrahedron Letters* **29**: 5343-5344.

23. Danilov LL, Druzhinina TN, Kalinchuk NA, Maltsev SD, Shibaev VN (1989) Polyprenyl phosphates: synthesis and structure-activity relationship for a biosynthetic system of Salmonella anatum O-specific polysaccharide. *Chem Phys Lipids* **51**: 191-203.

24. Danilov LL, Chojnacki T (1981) A Simple Procedure for Preparing Dolichyl Monophosphate by the Use of Poc13. *Febs Letters* **131**: 310-312.

25. Osborn JM, Cynkin MA, Gilber JM, Muller L, Singh M (1972) Synthesis of bacterial O-antigens. *Methods Enzymol* **28**: 583-601.

Mark M. Chen

Email: markchen@mit.edu Cell: (857) 753-5698 Work: (617) 253-0206

32 Vassar Street
MIT Building 18-544
Cambridge, MA 02139

Education

Massachusetts Institute of Technology — 2004-present
Ph.D. candidate in Biological Chemistry, expected graduation Feb 2009

University of British Columbia — 2000-2004
B.Sc. Double Major in Chemistry & Biochemistry, with honors

Research Experience

MIT Department of Chemistry, Supervisor Prof. Barbara Imperiali — 2004-present
-Ph.D. thesis involved a multidisciplinary approach, applying synthetic organic chemistry, solid phase peptide synthesis, cloning, recombinant expression, and enzyme kinetics to study protein glycosylation.
-Carried out a chemo-enzymatic synthesis of polyisoprene-linked carbohydrates donors and library of peptide acceptors to probe the substrate specificity of membrane bound enzymes *in vitro* from the N-linked glycosylation pathway of *Campylobacter jejuni.*

University of Leeds (UK) Biochemistry, Supervisor Prof. Sheena Radford — 2008 Fall
-Invited as a visiting scientist for a multidisciplinary collaborative project with MIT to investigate the effect of N-linked glycosylation on the kinetics of protein folding

UBC Department of Biochemistry, Supervisor Prof. Stephen Withers — 2003-2004
-B.Sc. thesis involved the use of mechanistic enzymology and medicinal chemistry to elucidate the function and mechanism of glycosidases of pharmaceutical interest.
-Designed and completed a 12 step synthesis of 7-deoxy-1,2-difluoro sialic acid as a mechanistic inhibitor of leech sialidase, as well as a 11 step synthesis of truncated difluoro sialic acid as an inhibitor of *Trypanosoma cruzi* trans-sialidase.

UBC Biotechnology Laboratories, Supervisor Dr. Robin Turner — 2002 Summer
-Conducted research in the area of biophysics and analytical chemistry, applying Raman laser spectroscopy and computer modeling methods for the development of an assay for the assessment of coronary atherosclerosis *in situ.*

UBC Department of Chemistry, Supervisor Prof. Laurel Schafer — 2001 Summer
-Conducted research in the area of organometallic chemistry, synthesizing and characterizing a small library of Lanthanide catalysts.

UBC Department of Chemistry, Supervisor Dr. Yoshikata Koga — 2001 Summer
-Conducted research in the area of physical and analytical chemistry, measuring the chemical potential of water + alcohol solution mixtures.

McDonald's Restaurant, Vancouver, B.C. Canada — 2000-2001
-Cashier; recognized for speed of service & courtesy towards customers

Summary of Laboratory Techniques

Protein: cloning, recombinant expression, purification, membrane protein handing, protein refolding, carbohydrates and protein glycosylation — 2003-present
Instruments: mass spectrometry, analytical ultracentrifugation, calorimetry
Chromatography: HPLC, LC, GC, normal phase, reverse phase, SDS-PAGE, gel filtration, ionic exchange, flash silica, thin layer
Synthesis: organic / medicinal chemistry, Fmoc solid phase peptide synthesis
Spectroscopy: fluorescence, UV-Vis, IR, NMR, circular dichroism, optical and radioactivity based assays
Supervision and training of junior lab members

Mark M. Chen Email: markchen@mit.edu Cell: (857) 753-5698 Work: (617) 253-0206

Publications

1. Reid CW, Stupak J, Chen MM, Imperiali B, Li J, Szymanski "*In vitro* metabolomic study of *N*-linked protein glycosylation." **Anal Chem** 14: 5458-75. — 2008 Summer
2. Chen MM, Weerapana E, Ciepichal E, Stupak J, Reid CW, Swiezewska E, Imperiali B "Polyisoprene specificity in the *Campylobacter jejuni N*-linked glycosylation pathway." **Biochemistry** 46: 14342-8. — 2007 Winter
3. Chen MM, Glover KJ, Imperiali B "From peptides to protein: comparative analysis of the substrate specificity of *N*-linked glycosylation in *C. jejuni*." **Biochemistry** 46: 5579-85. — 2007 Spring
4. Olivier NB, Chen MM, Behr JR, Imperiali B "*In vitro* biosynthesis of UDP-*N,N'*-diacetylbacillosamine by enzymes of the *Campylobacter jejuni* general protein glycosylation system." **Biochemistry** 45: 13659-69. — 2006 Winter
5. Glover KJ, Weerapana E, Chen MM, Imperiali B "Direct biochemical evidence for the utilization of UDP-bacillosamine by PglC, an essential glycosyl-1-phosphate transferase in the *Campylobacter jejuni N*-linked glycosylation pathway." **Biochemistry** 45: 5343-50. — 2006 Spring
6. Hackenberger CP, Chen MM, Imperiali B "Expression of *N*-terminal Cys-protein fragments using an intein refolding strategy." **Bioorg Med Chem** 14: 5043-8. — 2006 Summer
7. Weerapana E, Glover KJ, Chen MM, Imperiali B "Investigating bacterial *N*-linked glycosylation: synthesis and glycosyl acceptor activity of the undecaprenyl pyrophosphate-linked bacillosamine." **J Am Chem Soc** 127: 13766-7. — 2005 Fall
8. Hu J, Haynes CA, Wu AHY, Cheung CMW, Chen MM, Yee EGM, Ichioka T, Nishikawa K, Westh P, Koga Y "Chemical potential and concentration fluctuation in some aqueous akane-mono-ols at 25°C." **Can J Chem** 81: 141-9. — 2003 Winter

Presentations

Oral Presentation, **Northeastern Student Chemistry Research Conference** (Boston) Mark M. Chen, Barbara Imperiali, "Polyisoprenol Specificity in N-linked Glycosylation" — 2008 Spring

Oral Presentation, **MIT Department of Chemistry and Biology Joint Seminar** Mark M. Chen, Barbara Imperiali, "Polyisoprenol and Peptide Specificity of Bacterial Oligosaccharyltransferase PglB" — 2007 Winter

Poster Presentation, **National American Chemical Society Meeting** (Boston) Mark M. Chen, Barbara Imperiali, "*Campylobacter jejuni* Oligosaccharyl Transferase PglB Donor and Acceptor Specificity" — 2007 Summer

Oral Presentation, **Western Canadian Undergraduate Chemistry Conference** (Winnipeg) Mark M. Chen, Stephen G. Withers, "Synthesis of a Potential Competitive Inhibitor for *Trypanosomal* Trans-Sialidase" received **First Prize** in presentations — 2004 Spring

Oral Presentation, **UBC Multidisciplinary Undergraduate Research Conference** Mark M. Chen, Stephen G. Withers, "The Role of Sugars in Microbial Infections" — 2004 Spring

Oral Presentation, **National Undergraduate Chemistry Conference** (Ottawa) Mark M. Chen, Stephen G. Withers, "Synthesis of 7-deoxy-2,3-difluoro-N-acetylneuraminic acid as a Mechanism Based Inhibitor for Leech Sialidase" — 2003 Fall

Oral Presentation, **UBC Faculty of Science Student Research Showcase** Mark M. Chen, Robin F.B. Turner, Michael W. Blades, "The Assessment of Human Atherosclerosis with Raman Microscopy" — 2003 Spring

Mark M. Chen Email: markchen@mit.edu Cell: (857) 753-5698 Work: (617) 253-0206

Scholarships and Awards

MIT Chemistry Graduate Student Research Fellowship, *$64,000 per year*	2005- present
MIT Chemistry Graduate Student Teaching Fellowship, *$64,000*	2004-2005
NSERC Canadian Undergraduate Student Research Award, *$5000 per year* (The **Natural Science and Engineering Research Council** is the Canadian equivalent of the American National Institute of Health government funding agency. Each of the three awards was applied for separately.)	2001-2004
UBC Agnes and Gilbert Hooley Scholarship in Chemistry, *$1800*	2003 Summer
UBC Faculty of Medicine Summer Student Research Award, *$5000 declined*	2003 Summer

Teaching Experience

MIT 5.12 Organic Chemistry I lecture course teaching assistant -taught small group recitation to undergraduate students -prepared notes, problems, and lecture presentations to help students better understand the course material and prepare them for exams -received one of the highest TA evaluations from students in that class	2005 Spring
MIT 5.310 Laboratory Chemistry lab course teaching assistant -demonstrated and supervised multi-day lab experiments in general chemistry to a class of undergraduate students	2004 Fall

Leadership & Extracurricular Experience

MIT Chemistry Graduate Student Society - General Executive	2004-2008
MIT Graduate Student Society - Department of Chemistry Representative	2005-2008
MIT Sidney Pacific Graduate Dorm - Controller/Treasurer -managed and allocated an annual budget of approximately $60,000	2005-2007
Chinese Student Scholar Association, Boston Chapter -volleyball organizer and intramural team captain	2007-present
UBC Undergraduate Chemistry Society -'04 Academic Director, **'03 President**, '02 External Correspondent -Initiated and directed more than 25 academic and social events working with a team of 20 executives. Awarded the **2003 Student Chapters' Merit Award** from the Canadian Society for Chemistry, ranking #1 out of 73 university and college chapters.	2000-2004
UBC Biochemistry Physiology & Pharmacology Club -'04 Senior Biochemistry Dept Rep, '03 Treasurer, '02 General Executive -initiated among many events a very popular Professor Research Night which later became an annual tradition.	2001-2004
UBC Golden Key International Honor Society - General Executive	2001-2004
UBC Science Undergraduate Society - Dept of Biochemistry Representative	2002-2004
UBC Department of Chemistry - Faculty Curriculum Committee	2002-2003
Vancouver Board of Trade, Leaders of Tomorrow Program -chosen to participant as a mentee and event organizer	2003-2004

Languages

-English as first language
-conversational in Chinese (both Mandarin and Shanghainese dialects)

Mark M. Chen Email: markchen@mit.edu Cell: (857) 753-5698 Work: (617) 253-0206

References

Prof. Barbara Imperiali - MIT Dept of Chemistry & Dept of Biology — 2004-present
Ph.D. thesis supervisor overlooking research direction and progress
Professor for 5.55 Chemical Tools for Assessing Biological Function
Email: imper@mit.edu Phone: (617) 253-1838

Prof. Alexander Klibanov - MIT Dept of Chemistry & Dept of Bioengineering — 2004-present
chair of departmental committee in charge of evaluating PhD candidacy
Professor for 5.32 Biological Chemistry Tutorial
Email: klibanov@mit.edu Phone: (617) 253-3556

Prof. Stephen Withers - UBC Dept of Chemistry & Dept of Biochemistry — 2003-2004
B.Sc. thesis supervisor overlooking research direction and progress
Professor for Bioc403 Enzymology
Email: withers@chem.ubc.ca Phone: (604) 822-3402 Canada

Mr. Paul Stinson - President, CAPRA International Inc — 2003-present
mentor from the Vancouver Board of Trade, Leaders of Tomorrow Program
Email: paulstinson@capra-international.com
Phone: (778) 999-7488 Canada Phone: (86-136) 7185-1531 China